华章数学译丛

8

Matrix Analysis

矩阵分析

(美) Roger A. Horn 约翰斯·霍普金斯大学
Charles R. Johnson 威廉姆和玛丽学院 著

杨奇 译



机械工业出版社
China Machine Press

矩阵分析

"毫无疑问，对数值计算研究人员来说，本书是一本标准的参考书。"

——*Computing Reviews*

"不论对从事线性代数纯理论研究还是从事其应用研究的人员来说，本书都是一本必备的参考书。"

——*SIAM Review*

"这本书无疑会成为一本标准的教科书。"

——*American Scientist*

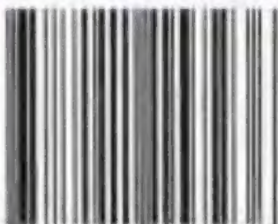
"总之，作者已经完成了一项杰出的工作，对线性代数和应用数学进行了精心组织的、内容全面广泛的综述，它既可以作为教科书，也可以作为参考书。对相关领域的每个人来说，本书都是必备的参考书。"

——*American Scientist*

矩阵理论作为一种基本的数学工具，在数学学科与其他科学技术领域（如数值分析、优化理论、微分方程、概率统计、系统工程等）都有广泛应用。电子计算机及计算技术的发展也为矩阵理论的应用开辟了更广阔的前景。因此，学习和掌握矩阵的基本理论和方法，对于理工科本科生和研究生来说是必不可少的。

本书论述了矩阵分析的经典方法和现代方法，包括了由于数学分析的需要而产生的线性代数中的论题，是解决实的和复的线性代数问题的重要方法之一。本书自1985年问世以来，受到越来越多的数学工作者和科技人员的好评和欢迎。时至今日，本书仍旧是一本十分有价值的数学著作。

Matrix Analysis





华章网站 http://www.hzbook.com
网上购书：www.china-pub.com
北京市西城区百万庄南街1号 100037
读者服务热线：(010)68995259，68995264
读者服务信箱：hzedu@hzbook.com

ISBN 7-111-15723-0/O・391 (课)
定价：45.00 元

华 章 数 学 译 丛

8

Matrix Analysis

矩阵分析

Roger A. Horn
约翰斯·霍普金斯大学
(美) Charles R. Johnson 著
威廉姆和玛丽学院
杨奇 译



机械工业出版社
China Machine Press

本书从数学分析的角度论述矩阵分析的经典方法和现代方法，取材新，有一定的深度，并给出在多元微积分、复分析、微分方程、最优化、逼近理论中的许多重要应用. 主要内容包括：特征值、特征向量和相似性，酉等价和正规矩阵，标准形，Hermite矩阵和对称矩阵，向量范数和矩阵范数，特征值的估计和扰动，正定矩阵，非负矩阵.

本书可作为工程、统计、经济学等专业的研究生教材或数学专业高年级本科生教材，也可以作为数学工作者和科技人员的参考书.

Roger A. Horn and Charles R. Johnson: Matrix Analysis(ISBN 0-521-38632-2).

Originally published by Cambridge University Press in 1985. Reprinted with corrections 1987, reprinted 1988, 1990, 1991, 1992, 1993, 1994, 1996, 1999.

本书版权登记号：图字：01-2004-6505

图书在版编目（CIP）数据

矩阵分析/(美)合恩(Horn, R. A.)等著；杨奇译. —北京：机械工业出版社，2005.4

(华章数学译丛)

书名原文：Matrix Analysis

ISBN 7-111-15723-0

Ⅰ. 矩…　Ⅱ. ①合…②杨…　Ⅲ. 矩阵分析　Ⅳ. O151.21

中国版本图书馆CIP数据核字(2004)第128206号

机械工业出版社(北京市西城区百万庄大街22号　邮政编码　100037)

责任编辑：迟振春

北京瑞德印刷有限公司印刷·新华书店北京发行所发行

2005年4月第1版第1次印刷

787mm×1092mm 1/16・25.75印张

印数：0 001-4 000册

定价：45.00元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换

本社购书热线：(010)68326294

译 者 序

由著名美国数学家合恩(Horn, R. A.)教授和约翰逊(Johnson, C. R.)教授合著的《Matrix Analysis》是一本不可多得的优秀教材. 它从数学分析的角度论述矩阵分析的经典结果和现代结果. 原书逻辑清晰, 结构严谨, 取材广泛, 内容新颖, 注重应用. 原书不仅重视矩阵分析的教学, 而且重视科学研究能力的培养. 在大部分章节末尾, 原书对所讨论的论题都提供了有关的参考文献, 并且独具匠心地编排了许多具有探究性和启发性的习题, 引导读者提高描述和解决数学问题的能力. 原书自 1985 年问世以来, 多次再版和重印, 受到越来越多的数学工作者和科技人员的好评和欢迎. 时至今日, 原书仍旧是一本十分有价值的名著, 对其赞美之词不绝于耳; 多家有影响的学术刊物高度评价作者的杰出工作以及原书在大学数学教育以及科学研究中的重要地位.

1988 年, 译者将原书译成中文介绍给我国读者. 受当时条件的限制, 译者忍痛割舍了原书部分内容、供进一步阅读的参考文献和参考书目以及部分习题; 而且译本印数较少, 未能满足广大读者的需要. 所幸的是, 机械工业出版社顺应了读者的要求, 决定重新出版该译本, 借再版的机会, 除了勘误以外, 译者还将割舍的内容全部补齐, 并且根据原书 1987 年修订本的 1999 年重印版作了修正, 使译本趋于完整且更具参考价值.

关于原书的撰写意图、特色、内容概要以及使用该书的建议等可参阅前言, 这里不再赘述.

书后"索引"中的英文名词术语仍按英文字母顺序排列, 再附加中文注释, 其页码仍指原书页码, 以便读者查阅有关的内容.

由于译者水平有限, 不当之处在所难免, 恳请读者惠予指正.

译 者

2004 年 11 月

前　言

线性代数和矩阵理论很久以来就已经是各数学学科的基本工具，而且就其本身的研究来说它们也是富有创造性的领域. 在本书中，给出了矩阵分析的经典结果和现代结果，这些结果已被证明对应用数学是重要的. 本书可作为大学本科生或研究生的教材，也可以作为各种不同读者的自学参考书. 我们假定读者已学过一个学期的基础线性代数课程，并具备初步的分析知识. 本书从讨论特征值和特征向量开始，所以不要求读者已熟悉这些概念.

除了在基础线性代数教程中会涉及到有关矩阵的内容以外，实际上，任何涉及数学的领域(包括微分方程、概率统计、最优化；理论经济学及应用经济学、工程学或运筹学等等)都需要矩阵的知识. 然而，直到最近，很多必需的内容才零星地(或者根本没有)出现在大学本科生或研究生的课程中. 随着应用数学越来越重要，有相当多的课程专门研究高等矩阵理论，这就迫切需要有一本能广泛选择各种论题的教材，并且它还能对有关的论题提供现代参考资料.

虽然有不少深受读者喜爱的矩阵理论方面的经典著作，但它们既不适合于一般的课堂教学，也不适合于系统的自学. 某些传统的参考书缺乏习题、不注重应用和启发诱导；索引不完善；处理方法陈旧. 而较现代的书又往往趋向于作为基础教材，或趋向于作为讨论专门化论题的专著. 我们的目标是撰写一本能对范围广泛的各种论题进行有效而现代处理的书.

矩阵分析的一个意图是，它要包括由于数学分析(例如，多元微积分、复变量、微分方程、最优化和逼近理论等)的需要而产生的线性代数中的论题. 矩阵分析的另一个意图是，它是解决实的和复的线性代数问题的一种方法，这种方法果断地采用诸如极限、连续和幂级数这些来自分析的概念，这些概念有时比纯代数方法更为有效或更为自然. 矩阵分析的这两个出发点影响了对本书中所述论题的选择和讨论. 我们认为采用术语*矩阵分析*比*线性代数*更能准确地反映该领域的广泛内容和研究方法.

为了复习和便于查阅，本书第 0 章介绍了基础线性代数所必需的内容，另外还包括一些有用的资料. 第 1 章、第 2 章和第 3 章涉及大部分核心内容，这些内容可能包括在线性代数或矩阵理论的任何选修课程中，主要有：特征值、特征向量和相似性；酉相似、Schur 三角化及其推论、正规矩阵；标准形和包括 Jordan 标准形在内的各种分解、LU 分解、QR 分解和友矩阵. 除此以外，以下各章大体上都是独立展开的，而且以一定的深度讨论一个主要论题：

Hermite 矩阵和复对称矩阵(第 4 章). 该章着重讨论了关于研究 Hermite 矩阵的特征值的变分法，其中包括优化概念的引入.

向量范数和矩阵范数(第 5 章). 它们对于数值线性代数算法的误差分析，对于研究矩阵幂级数和迭代过程是必不可少的. 该章较详细地讨论了范数的代数性质、几何性质和分析性质，并且仔细地区分了与矩阵范数的次乘性公理有关的和无关的矩阵范数结果.

特征值的估计和扰动(第 6 章). 这是针对一般矩阵(不一定是 Hermite 矩阵)来讨论的，并且对于许多应用是很重要的. 该章详细讨论了 Geršgorin 区域理论以及它的某些现代的改进，同时还给出了有关的图论概念.

正定矩阵(第 7 章). 该章详细地考察了正定矩阵及其包括不等式在内的各种应用，讨论了极分解和奇异值分解，同时还讨论了对矩阵逼近问题的应用.

非负矩阵(第 8 章). 它们起因于其中必然会出现非负量的许多应用(如概率论、经济学和工程技术等)，其引人瞩目的理论反映了这些应用. 在运用范数的基础上，该章初步讨论了关于非负矩阵、正矩阵、素矩阵和不可约矩阵的理论.

在本书的姊妹篇中，将讨论其他一些同样值得注意的论题：值域及推广；惯性、稳定矩阵、M-矩阵和相关特殊类；矩阵方程、Kronecker 乘积和 Hadamard 乘积；可以把函数和矩阵联系起来的各种方法.

本书为一个学期或两个学期的课程提供了基本教材. 根据特殊课程的需要，教师可对本书的内容进行合理的选择. 一般应当包括第 1 章、第 2 章和第 3 章的大部分内容以及第 4 章和第 5 章中有关 Hermite 矩阵和范数的内容.

本书大多数章节包括了一些比较专门化的或非传统的材料. 例如，第 2 章不仅包括一个矩阵的酉三角化的 Schur 基本定理，而且还讨论了矩阵族的同时三角化. 在关于酉等价一节中，在介绍通常的内容之后，还讨论了两个矩阵是酉等价的迹条件. 第 4 章中关于复对称矩阵的讨论是对照推导 Hermite 矩阵的经典理论给出的. 一个论题的基本观点出现在每一章的前几节中，而更为详细的讨论则出现在各节的末尾或以后各节中. 这样做的优点是可以依次提出各种论题，从而提高了本书作为参考书的使用价值，并为教师提供了广泛的选择余地.

本书的许多结果(或者经推广后)适合于其他域上的矩阵，或者可以在一些更广泛的代数系统中讨论这些结果. 不过，我们有意将定义域限制在实数域或复数域，这样就能够使用经典分析的熟悉方法以及正规的代数方法.

虽然我们一般考虑复矩阵，但大多数例子只限于实矩阵，并且不需要较深的复分析知识. 熟悉复数的算术对于理解矩阵分析是必不可少的，其熟悉程度需达到附录中所规定的水平. 其余几个简短的附录包括几个次要的而又必需的论题，例如，Weierstrass 定理和凸性.

在本书中，给出了许多练习和习题，为进一步理解主要论题及其推论奠定了基础. 练习作为每节的论述部分而出现，它们一般都很基本，直接用于加深对概念的理解，建议读者充分选做这些练习. 习题安排在每节的末尾(没有特定的顺序)，其中涉及一系列难题和典型题(既有证明题，也有计算题)，可以用来扩展论题，提出特殊见解，或者为主要论题提供另外的证明. 书中对较难的习题给出了重要的提示，有些习题的结果与其他习题或正文中的结果有关. 要特别强调的是，尽心竭力去完成练习和钻研习题对学习本书是至关重要的.

虽然本书不是一本有关应用方面的书，但为了激发学习兴趣，在每一章的开始都介绍几个应用来引人本章的论题.

如果想查阅某个论题的其他处理方式以及与之有关的资料，可参看"参考文献"所列著作. 在正文中用一个简明的助记码引用其中的著作；例如，Jones 和 Smith 的著作记作[JSm]. 这些助记码和全部引文在"参考文献"中按作者姓名的字母顺序排列.

"参考文献"并不详尽. 在一本有多个一般论题的书中，限于篇幅，在正文中我们尽可能压缩所引用文献的数量. 在大多数章节的末尾，我们列出了经过挑选的文献(例如明显用到了的

那些文献)，另外附带一个简短的讨论，但没有刻意去收集有关经典结果的历史资料. 众多的文献是由我们已参考的较专业的书籍中提供的. 读者还应当了解涉及矩阵分析内容的概括性现代文献资料. 例如，*KWIC Index for Numerical Linear Algebra*[CaLe]以及 *Mathematical Reviews* 的 15 节和 65 节.

感谢为本书提供有益建议的同事和学生，他们对本书的课堂笔记和手稿提出了修改意见. 他们是：Wayne Barrett，Leroy Beasley，Bryan Cain，David Carlson，Dipa Choudhury，Risana Chowdhury，Yoo Pyo Hong，Dmitry Krass，Dale Olesky，Stephen Pierce，Leiba Rodman，Pauline van den Driessche.

R. A. H.

C. R. J.

符号表

$\mathbf{R}$	实数域
$\mathbf{R}^n$	实 n 维列向量空间，$M_{n,1}(\mathbf{R})$
$\mathbf{C}$	复数域
$\mathbf{C}^n$	复 n 维列向量空间，$M_{n,1}(\mathbf{C})$
$\mathbf{F}$	一个域(通常指 $\mathbf{R}$ 或 $\mathbf{C}$)
$\mathbf{F}^n$	(域 $\mathbf{F}$ 上)分量取自 $\mathbf{F}$ 的 n 维列向量空间，$M_{n,1}(\mathbf{F})$
$M_{m,n}(\mathbf{F})$	元素取自 $\mathbf{F}$ 的 $m\times n$ 矩阵的集合
$M_{m,n}$	$m\times n$ 复矩阵的集合，$M_{m,n}(\mathbf{C})$
M_n	$n\times n$ 复矩阵的集合，$M_{n,n}(\mathbf{C})$
A，B，C，等等	矩阵；$A=[a_{ij}]\in M_{m,n}(\mathbf{F})$
x，y，z，等等	列向量；$x=[x_i]\in\mathbf{F}^n$
I	$M_n(\mathbf{F})$中的单位矩阵
0	零纯量，零向量或零矩阵
$\overline{A}$	$A\in M_{m,n}(\mathbf{C})$的各元素取复共轭的矩阵
A^T	$A\in M_{m,n}(\mathbf{F})$的转置
A^*	$A\in M_{m,n}(\mathbf{C})$的 Hermite 伴随，$\overline{A}^T$
A^{-1}	非奇异矩阵 $A\in M_n(\mathbf{F})$的逆
$A^{1/2}$	半正定矩阵 $A\in M_n$ 的唯一半正定平方根
$\|A\|$	$A\in M_{m,n}$的各元素取绝对值的矩阵
$A^\dagger$	$A\in M_{m,n}$的 Moore-Penrose 广义逆
$\operatorname{adj} A$	$A\in M_n(\mathbf{F})$的经典伴随(转置伴随)
$\mathscr{B}$	向量空间的一个基
e_i	(通常指)$\mathbf{F}^n$ 中第 i 个标准基向量
$[v]_{\mathscr{B}}$	向量 v 的 $\mathscr{B}$ 坐标表示
${}_{\mathscr{B}_2}[T]_{\mathscr{B}_1}$	线性变换 T 的 $\mathscr{B}_1$-$\mathscr{B}_2$ 基表示
$\binom{n}{k}$	二项式系数，$n!/[k!(n-k)!]$
$p_A(t)$	$A\in M_n(\mathbf{F})$的特征多项式
$\kappa(A)$	非奇异矩阵 $A\in M_n$(关于逆相应于给定的矩阵范数)的条件数
$\det A$	$A\in M_n(\mathbf{F})$的行列式
$\oplus$	直和
$\Gamma(A)$	$A\in M_n(\mathbf{F})$的有向图
$\\|\cdot\\|^D$	向量范数 $\\|\cdot\\|$ 的对偶范数

$f^D(\cdot)$	准范数 $f(\cdot)$的对偶范数
λ	(通常指)$A\in M_n$ 的特征值
$\{\lambda_i(A)\}$	$A\in M_n$ 的特征值的集合(谱)；如果 A 是 Hermite 矩阵，通常指 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$
$n!$	阶乘，$n(n-1)(n-2)\cdots2\cdot1$
$G(A)$	$A\in M_n$ 的 Geršgorin 区域
$GL(n, F)$	$M_n(\mathbf{F})$中的非奇异矩阵构成的群
$A\circ B$	A，$B\in M_{m,n}(\mathbf{F})$的 Hadamard 乘积
$\gamma(A)$	素矩阵 $A\in M_n$ 的本原指标
$M(A)$	$A\in M_{m,n}(\mathbf{F})$的指标矩阵
$J_k(\lambda)$	具有特征值 λ 的 k 阶 Jordan 块
$\otimes$	Kronecker(张量)乘积
$q_A(t)$	$A\in M_n(\mathbf{F})$的极小多项式
$\lVert\cdot\rVert_1$	$\mathbf{C}^n$ 的 l_1(和)范数；M_n 上的 l_1 矩阵范数
$\lVert\cdot\rVert_2$	$\mathbf{C}^n$ 的 l_2(Euclid)范数；M_n 上的 l_2(Frobenius)矩阵范数
$\lVert\cdot\rVert_\infty$	$\mathbf{C}^n$ 上的 l_∞(极大)范数；M_n 上的 l_∞ 向量范数
$\lVert\cdot\rVert_p$	$\mathbf{C}^n$ 上的 l_p 范数
$\lVert\lvert\cdot\rvert\rVert_1$	M_n 上的极大列和矩阵范数
$\lVert\lvert\cdot\rvert\rVert_2$	M_n 上的谱矩阵范数
$\lVert\lvert\cdot\rvert\rVert_\infty$	M_n 的极大行和矩阵范数
$r(A)$	(通常指)$A\in M_n$ 的数值半径
$\perp$	正交补
per A	$A\in M_n(\mathbf{F})$的积和式
rank A	$A\in M_{m,n}(\mathbf{F})$的秩
sgn	排列的正负号函数
$\{\sigma_i(A)\}$	$A\in M_{m,n}$的奇异值的集合；通常取 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_{\min\{m,n\}}\geqslant0$
$\sigma_1(A)$	$A\in M_{m,n}$的最大奇异值，$\lVert\lvert A\rvert\rVert_2$
Span S	向量空间的子集 S 的张成
$\rho(A)$	$A\in M_n$ 的谱半径
$\sigma(A)$	$A\in M_n$ 的谱(特征值的集合)
$A(\alpha, \beta)$	由指标集 α，β 确定的 $A\in M_{m,n}(\mathbf{F})$的子矩阵
tr A	$A\in M_n(\mathbf{F})$的迹

目　录

译者序

前言

符号表

第 0 章　复习及其他 …… 1

0.0　导引 …… 1

0.1　向量空间 …… 1

0.2　矩阵 …… 3

0.3　行列式 …… 5

0.4　秩 …… 8

0.5　非奇异性 …… 9

0.6　普通内积 …… 10

0.7　分块矩阵 …… 11

0.8　行列式(续) …… 13

0.9　矩阵的特殊形式 …… 16

0.10　基的变换 …… 21

第 1 章　特征值、特征向量和相似性 …… 25

1.0　导引 …… 25

1.1　特征值-特征向量方程 …… 26

1.2　特征多项式 …… 28

1.3　相似性 …… 32

1.4　特征向量 …… 41

第 2 章　酉等价和正规矩阵 …… 47

2.0　导引 …… 47

2.1　酉矩阵 …… 47

2.2　酉等价 …… 52

2.3　Schur 酉三角化定理 …… 57

2.4　Schur 定理的若干推论 …… 61

2.5　正规矩阵 …… 71

2.6　*QR* 分解和 *QR* 算法 …… 80

第 3 章　标准形 …… 85

3.0　导引 …… 85

3.1　Jordan 标准形：一个证明 …… 86

3.2　Jordan 标准形：若干论断和应用 …… 92

3.3　多项式和矩阵：极小多项式 …… 101

3.4　其他标准形和分解 …… 107

3.5　三角分解 …… 113

第 4 章　Hermite 矩阵和对称矩阵 …… 119

4.0　导引 …… 119

4.1　Hermite 矩阵的定义、性质和特征 …… 120

4.2　Hermite 矩阵的特征值的变分特征 …… 125

4.3　变分特征的某些应用 …… 130

4.4　复对称矩阵 …… 144

4.5　Hermite 矩阵、对称矩阵的相合与同时对角化 …… 156

4.6　合相似和合对角化 …… 174

第 5 章　向量范数和矩阵范数 …… 183

5.0　导引 …… 183

5.1　向量范数和内积的定义性质 …… 184

5.2　向量范数的例子 …… 188

5.3　向量范数的代数性质 …… 190

5.4　向量范数的分析性质 …… 191

5.5　向量范数的几何性质 …… 200

5.6　矩阵范数 …… 205

5.7　关于矩阵的向量范数 …… 227

5.8　矩阵的逆和线性方程组的解的误差 …… 238

第 6 章　特征值的估计和扰动 …… 245

6.0　导引 …… 245

6.1　Geršgorin 圆盘 …… 245

6.2　Geršgorin 圆盘——更细致的讨论 …… 252

6.3　扰动定理 …… 260

6.4　其他包含区域 …… 269

第 7 章　正定矩阵 …… 279

7.0　导引 …… 279

7.1　定义和性质 …… 282

7.2　正定矩阵的特征 …… 286

7.3　极形式和奇异值分解 …… 292

7.4 奇异值分解的例子和应用 …………………… 302

7.5 Schur 乘积定理 ………………………………… 323

7.6 相合：乘积和同时对角化 …………………… 329

7.7 半正定次序关系 ………………………………… 332

7.8 关于正定矩阵的不等式 ………………………… 337

第 8 章 非负矩阵 ………………………………… 315

8.0 导引 ……………………………………………… 315

8.1 非负矩阵——不等式及其推广 ……………… 317

8.2 正矩阵 …………………………………………… 351

8.3 非负矩阵 ………………………………………… 356

8.4 不可约非负矩阵 ………………………………… 359

8.5 素矩阵 …………………………………………… 364

8.6 一般极限定理 …………………………………… 370

8.7 随机矩阵和双随机矩阵 ……………………… 372

附录 A 复数 ………………………………………… 375

附录 B 凸集和凸函数 ……………………………… 377

附录 C 代数基本定理 ……………………………… 380

附录 D 多项式的零点对其系数的连续依赖性 …… 381

附录 E Weierstrass 定理 ………………………… 382

参考文献 …………………………………………… 383

索引 ………………………………………………… 387

第0章　复习及其他

0.0　导引

本章的目的是简要地编撰许多有用概念或结果，不加证明，其中许多概念和结果直接或间接地为涉及本书主要章节的内容奠定了基础．所编写内容的大部分可能以某种形式包括在线性代数的基础教程中，另外，还选编了一些有用的资料，这些资料往往不易在其他地方找到，或者不宜安排在以后的章节中．这样，在开始学习本书之前，这一章可作为读者的复习提要，必要时也为读者参阅有关资料提供方便．本章还规定了一些基本符号，并给出一些定义；为此，了解本章是有益的．这里，要求读者已经熟悉线性代数的基本概念和矩阵的基本运算，如矩阵的乘法和加法．

0.1　向量空间

在本书的论述中，虽然一般是含蓄地述及向量空间，但是，向量空间是矩阵理论的基本结构．

0.1.1 纯量域　构成向量空间的基础是域，或者是具有乘法的纯量集．对于实际应用来说，在通常的加法和乘法运算下，基域几乎总是实数域 **R** 或复数域 **C**(见附录 A)．但是，它也可能是有理数域，也可能是关于一个特定素数的整数同余类域或一些其他的域．当未指明是哪种域时，就用符号 **F** 表示域．为了验证一个纯量集是域，它必须在两个指定的二元运算("加法"和"乘法")下封闭；两个运算必须满足结合律和交换律，且在该集合中各有一个单位元；对于所有的元素，在该集合中必须有关于加法运算的逆元素，并且对于除加法单位元(0)以外的所有元素，在该集合中有关于乘法运算的逆元素；同时，乘法运算对加法运算必须满足分配律．

[1]

0.1.2 向量空间　域 **F** 上的向量空间是一些对象(称为向量)的集合 V，它在一个二元运算(加法)下封闭，这个运算是结合的和交换的，在集合 V 中有一个单位元("0")，且有加法逆元．该集合对用纯量域 **F** 的元素左乘向量的运算也是封闭的，且有性质：对所有的 a，$b\in\mathbf{F}$，以及所有的 x、$y\in V$，有 $a(x+y)=ax+ay$，$(a+b)x=ax+bx$，$a(bx)=(ab)x$，以及对乘法单位元 $e\in\mathbf{F}$，有 $ex=x$．

对于给定的域 **F**，分量取自 **F** 的 n 元组的集合 $\mathbf{F}^n$(n 是整数)，在通常的运算(在 $\mathbf{F}^n$ 中按分量相加)下构成 **F** 上的一个向量空间．特别地，$\mathbf{R}^n$ 和 $\mathbf{C}^n$ 是本书的基本向量空间．具有实系数或复系数的(不超过某一指定次数的或任意次数的)多项式集合，以及区间$[a, b]\subset\mathbf{R}$上的实值或复值连续函数，或任意函数的集合也是(**R** 上或 **C** 上)向量空间的例子．当然，在有限维空间 $\mathbf{R}^n$ 与由$[0, 1]$上的实值连续函数组成的无限维向量空间之间，有着本质的差别．

0.1.3 子空间和张成　向量空间 V 的子空间 U 是 V 的非空子集，它自身正好是同一个纯量域上的向量空间．例如$[a, b, 0]^T$：a、$b\in\mathbf{R}$ 是 $\mathbf{R}^3$ 的子空间．通常，我们用某种关系来定义向

量空间 V 的子空间，如此得到的由 V 中部分元素组成的集合关于 V 中的加法是封闭的．例如，$\mathbf{R}^3$ 中最后一个分量是0的所有元素组成的集合．一般认为，把所得到的集合看成一个子空间比自身看成一个向量空间更有用．在任何情况下，两个子空间的交还是子空间．

如果 S 是向量空间 V 的子集，S 的张成是集合 $\mathrm{Span}S=\{a_1v_1+a_2v_2+\cdots+a_kv_k:a_1,a_2,\cdots,a_k\in\mathbf{F},v_1,v_2,\cdots,v_k\in S,k=1,2,\cdots\}$．注意，即使 S 不是子空间，$\mathrm{Span}\ S$ 总是子空间．如果 $\mathrm{Span}\ S=V$，就称 S 张成向量空间 V． [2]

0.1.4 线性相关和线性无关 称一个向量空间中的向量组 $\{x_1,x_2,\cdots,x_k\}$ 线性相关，指的是在纯量基域 $\mathbf{F}$ 中存在不全为0的系数 $a_1,a_2,\cdots,a_k$，使得

$$a_1x_1+a_2x_2+\cdots+a_kx_k=0$$

或等价地，某一向量 x_i 是其余向量的线性组合，其中系数取自 $\mathbf{F}$．例如，$\{[1,2,3]^T,[1,0,-1]^T,[2,2,2]^T\}$ 是 $\mathbf{R}^3$ 的中线性相关组．V 中的一个子集在 $\mathbf{F}$ 上不线性相关，就说它线性无关．例如，$\{[1,2,3]^T,[1,0,-1]^T\}$ 是 $\mathbf{R}^3$ 中的线性无关组．重要的是要注意这两个概念实质上与向量组有关．线性无关的任一非空子集线性无关；$\{0\}$ 是线性相关组；因而，包含0向量的任一集合线性相关．可能一个向量集线性相关，而它的任一真子集是线性无关的．

0.1.5 基 设 S 是向量空间 V 的子集．如果 V 的每个元素可以表示成 S 的诸元素（具有纯量基域中的系数）的线性组合，就称 S 张成 V．例如，$\{[1,0,0]^T,[0,1,0]^T,[0,0,1]^T,[1,0,1]^T\}$ 在 $\mathbf{R}$ 上张成 $\mathbf{R}^3$（或在 $\mathbf{C}$ 上张成 $\mathbf{C}^3$）．张成向量空间 V 的一个线性无关组为 V 的一个基．基虽然不是唯一的，但是基很有用，V 的每个元素能且只能用一种方式由基来表示，且在该基中再添加任何一个元素或从该基中去掉任一元素，上述性质不再成立．V 中的无关组是 V 的一个基，当且仅当没有真包含它的无关组．张成 V 的一个集合是 V 的一个基，当且仅当它没有真子集仍张成 V．每个向量空间总有一个基．

0.1.6 扩充成一个基 向量空间 V 中的任何线性无关组都可以扩充为 V 的一个基，也就是说，给定 V 中的线性无关组 $\{x_1,x_2,\cdots,x_k\}$，存在另外的向量 $x_{k+1},\cdots,x_n,\cdots\in V$，使得 $\{x_1,x_2,\cdots,x_n,\cdots\}$ 是 V 的一个基，把一个已知的无关组扩充为一个基，当然不是唯一的［例如，可以把第三个分量是非零的任一向量添加到无关组 $\{[1,0,0]^T,[0,1,0]^T\}$ 中，使得到 $\mathbf{R}^3$ 的一个基］．由 $[0,1]$ 上实值连续函数组成的实向量空间 $C[0,1]$ 的例子说明，一般地，一个基未必有限；由单项式 $\{1,x,x^2,x^3,\cdots\}$ 组成的无限集是 $C[0,1]$ 中的无关组．

[3] **0.1.7 维数** 如果向量空间 V 的某个基包含有限个元素，那么所有的基有相同的元素个数，并且称这个公共的数为向量空间的维数．这时，就说 V 是有限维的，否则，就说 V 是无限维的．无限维的情形（例如，$C[0,1]$），在任意两个基的元素之间存在一个一一对应．实向量空间 $\mathbf{R}^n$ 有维数 n．向量空间 $\mathbf{C}^n$ 在域 $\mathbf{C}$ 上有维数 n，但在域 $\mathbf{R}$ 上有维数 $2n$．有时称基 $\{e_1,e_2,\cdots,e_n\}$ 为 $\mathbf{R}^n$ 或 $\mathbf{C}^n$ 的标准基，其中 e_i 的第 i 个分量是1，其余分量是0．

0.1.8 同构 如果 U 和 V 是同一个纯量域 $\mathbf{F}$ 上的向量空间，且 $f:U\rightarrow V$ 是可逆函数，使得对所有的 $x,y\in U$ 和所有的 $a,b\in\mathbf{F}$ 有 $f(ax+by)=af(x)+bf(y)$，则 f 是一个同构，且称 U 和 V 同构（"结构相同"）．同一个域上的两个有限维向量空间同构，当且仅当它们有相同的维

数：于是，域 F 上的任一 n 维向量空间同构于 $\mathbf{F}^n$. 因此，任一 n 维实向量空间同构于 $\mathbf{R}^n$，而任一 n 维复向量空间同构于 $\mathbf{C}^n$. 特别是，如果 V 是域 F 上的 n 维向量空间，具有一个给定的基 $\mathscr{B}=\{x_1, \cdots, x_n\}$，那么，因为任一元素 $x\in V$ 可以唯一地写成 $x=a_1x_1+\cdots+a_nx_n$，$a_i\in \mathbf{F}$，$i=1, 2, \cdots, n$，于是，相对于这个基，可以把 x 对应于 n 元组 $[x]_{\mathscr{B}}=[a_1, \cdots, a_n]^T$. 对于任一个基 $\mathscr{B}$，映射 $x\to[x]_{\mathscr{B}}$ 是 V 与 $\mathbf{F}^n$ 之间的一个同构.

0.2 矩阵

这里所研究的对象可以用两种重要的方式来考察：一是把它看成纯量的矩形阵列，一是给每个空间指定一个基，然后把它看作两个向量空间之间的线性变换.

0.2.1 矩形阵列 一个矩阵是由域 F 中若干个纯量组成的一个 $m\times n$ 阵列. 如果 $m=n$，就称矩阵是方阵. F 上的所有 $m\times n$ 的矩阵集合用 $M_{m,n}(\mathbf{F})$ 来表示，而 $M_{n,n}(\mathbf{F})$ 简记为 $M_n(\mathbf{F})$. 最常见的情形是 $\mathbf{F}=\mathbf{C}$（复数域）. 还把 $M_n(\mathbf{C})$ 简记为 M_n，$M_{m,n}(\mathbf{C})$ 简记为 $M_{m,n}$. 通常用大写字母来表示矩阵. 例如，如果

$$A=\begin{bmatrix} 2 & -\frac{3}{2} & 0 \\ -1 & \pi & 4 \end{bmatrix},$$

那么 $A\in M_{2,3}(\mathbf{R})$. 一个给定矩阵的子矩阵是位于该矩阵的一些指定的行和列的矩形阵列. 例如，$[\pi, 4]$ 是上述 A 的子矩阵（位于第 2 行，第 2 列，第 3 列）.

4

0.2.2 线性变换 设 U 和 V 分别是同一个纯量域 F 上的 n 维向量空间和 m 维向量空间；设 $\mathscr{B}_U$ 和 $\mathscr{B}_V$ 分别是 U 和 V 的基. 我们可以分别用同构 $x\to[x]_{\mathscr{B}_U}$ 和 $y\to[y]_{\mathscr{B}_V}$ 把 U 和 V 中的向量表示成 F 上的 n 元组和 m 元组. 一个线性变换是一个函数 $T: U\to V$，使得对于任意纯量 a_1 和 a_2，以及向量 x_1 和 x_2，都有 $T(a_1x_1+a_2x_2)=a_1T(x_1)+a_2T(x_2)$. 一个矩阵 $A\in M_{m,n}(\mathbf{F})$ 可以用下述方式对应于一个线性变换 $T: U\to V$：向量 $y=T(x)$ 当且仅当 $[y]_{\mathscr{B}_V}=A[x]_{\mathscr{B}_U}$. 这时就称矩阵 A 表示线性变换 T（关于基 $\mathscr{B}_U$ 和 $\mathscr{B}_V$）；表示矩阵 A 与基的选择有关. 在讨论矩阵时，要意识到是在讨论关于特别选定的基下的线性变换，但借助于什么基，一般不必明言.

0.2.3 与一个已知矩阵或线性变换相关联的向量空间 不失一般性，使 F 上的 n 维向量空间与 $\mathbf{F}^n$ 相对应，于是，就把 $A\in M_{m,n}(\mathbf{F})$ 看作从 $\mathbf{F}^n$ 到 $\mathbf{F}^m$ 的线性变换（同时也看作一个阵列）. 这样一个线性变换的定义域是 $\mathbf{F}^n$；它的值域是 $\{y\in \mathbf{F}^m: y=Ax$，对所有 $x\in \mathbf{F}^n\}$. A 的零空间是 $\{x\in \mathbf{F}^n: Ax=0\}$. A 的值域是 $\mathbf{F}^m$ 的子空间，而 A 的零空间是 $\mathbf{F}^n$ 的子空间. 关于这两个子空间的关系式是：

$$n=A \text{ 的零空间的维数}+A \text{ 的值域的维数}.$$

0.2.4 矩阵运算 矩阵加法定义为两个同维阵列按对应元相加，并且用 +（"$A+B$"）表示. 它对应线性变换的加法（关于相同的基），且继承了从纯量域来的交换性和结合性. 零矩阵（所有元全为 0 的矩阵）是矩阵加法的单位元，并且 $M_{m,n}(\mathbf{F})$ 自身也是 F 上的向量空间. 按通常方式定义的矩阵乘法用 AB 来表示，它与线性变换的复合相对应. 这样，只有当 $A\in M_{m,n}(\mathbf{F})$，$B\in$

[5] $M_{p,q}(\mathbf{F})$，且 $p=n$ 时，它才有定义；它是结合的，一般是不交换的．例如，

$$\begin{bmatrix}1 & 0\\0 & 2\end{bmatrix}\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\neq\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 2\end{bmatrix}.$$

但是，当把矩阵限制在 $M_n(\mathbf{F})$ 的某些有研究价值的子集时，它可以是交换的．矩阵乘法有一个单位元，即形如

$$I=\begin{bmatrix}1 & & & 0\\ & 1 & & \\ & & \ddots & \\ 0 & & & 1\end{bmatrix}$$

的矩阵 $I\in M_n(\mathbf{F})$．这个矩阵以及它的所有纯量倍数（称为纯量矩阵）与 $M_n(\mathbf{F})$ 中的所有其他矩阵都可交换，并且只有纯量矩阵具有这一性质．矩阵乘法对于矩阵加法是分配的．

这里需要指出，我们总是用符号 0 表示以下各种术语：零纯量、零向量（所有分量都等于零纯量的向量）和零矩阵（所有的元都等于零纯量）．一般地，上下文将明确它是哪种情形，因而不会引起混淆．我们还用符号 I 表示任意阶数的单位矩阵．如果可能引起混淆，就指明其阶数．

0.2.5 转置与 Hermite 伴随 如果 $A=[a_{ij}]\in M_{m,n}(\mathbf{F})$，$A$ 的转置，记作 A^T，是 $M_{n,m}(\mathbf{F})$ 中的一个矩阵，它的元是 a_{ji}；即将原矩阵行与列调换，反之亦然．例如，

$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\end{bmatrix}^T=\begin{bmatrix}1 & 4\\2 & 5\\3 & 6\end{bmatrix}.$$

显然，$(A^T)^T=A$．$A\in M_{m,n}(\mathbf{C})$ 的 Hermite 伴随 A^* 定义为 $A^*=\bar{A}^T$，其中 $\bar{A}$ 表示按分量取共轭．例如，

$$\begin{bmatrix}1+i & 2 & i\\-3 & -2i\end{bmatrix}^*=\begin{bmatrix}1-i & -3\\2+i & 2i\end{bmatrix}.$$

转置和 Hermite 伴随［以及将在(0.5)中讨论的矩阵的逆］都服从倒序律：$[AB]^*=B^*A^*$ 和 $[AB]^T=B^TA^T$，当然要假定乘积有定义．对于乘积的共轭，不存在倒序：$\overline{AB}=\bar{A}\bar{B}$．如果 x，$y\in M_{n,1}=\mathbf{C}^n$，那么 y^*x 是纯量，并且它的 Hermite 伴随与它的复共轭相同；因此，$(y^*x)^*=\overline{y^*x}=x^*y=y^T\bar{x}$．

[6] **0.2.6 矩阵乘法的技巧** 这里，给出几个要反复用到的矩阵乘法的简单性质．

1. 如果 b_j 表示矩阵 B 的第 j 列，那么乘积 AB 的第 j 列正好是 Ab_j．

2. 如果 a_i 表示矩阵 A 的第 i 行，那么乘积 AB 的第 i 行正好是 a_iB．

解释一下，在乘积 AB 中，左乘以 A 是乘 B 的列，而右乘以 B 是乘 A 的行．对其中一个因子是对角矩阵的情形，在(0.9.1)中再讨论．

3. 如果 $A\in M_{m,n}(\mathbf{F})$，且 $x\in\mathbf{F}^n$，那么 Ax 是（以 x 的坐标为系数的）A 的各列的线性组合．

4. 如果 $A\in M_{m,n}(\mathbf{F})$，且 $y\in\mathbf{F}^m$，那么 y^TA 是（以 y 的坐标为系数的）A 的各行的线性组合．

0.3 行列式

只用一个数概括一种多变量现象，这在数学中常常很有用，其中行列式就是一例．行列式只对方阵 $A\in M_n(\mathbf{F})$ 有定义，并且它可以按两种重要的，完全不同的等价方式来描述．我们把 $A\in M_n(\mathbf{F})$ 的行列式记作 $\det A$.

0.3.1 Laplace 展开 对 $A=[a_{ij}]\in M_n(\mathbf{F})$ 的行列式可归纳定义如下．假设行列式在 $M_{n-1}(\mathbf{F})$ 上已定义，设 $A_{ij}\in M_{n-1}(\mathbf{F})$ 表示从 $A\in M_n(\mathbf{F})$ 中划去第 i 行和第 j 列后得到的子矩阵．于是，对所有的 $i\leqslant n$，$j\leqslant n$，

$$\sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij}=\sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij},$$

而这个相同的值就是 $\det A$．等式左边是依第 i 行关于诸子式的 Laplace 展开式，而右边是依第 j 列的 Laplace 展开式[见(0.7.1)]．对于任意选择的行和列，其中任一展开式都得到 A 的行列式．归纳过程从 1×1 矩阵开始，定义它的行列式为单个元的值．于是

[7]

$$\det[a_{11}]=a_{11},$$

$$\det\begin{bmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{bmatrix}=a_{11}a_{22}-a_{12}a_{21},$$

$$\det\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}=a_{11}a_{22}a_{33}-a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}-a_{13}a_{22}a_{31},$$

等等．显然，如果 $A\in M_n(\mathbf{C})$，则 $\det A^T=\det A$，且 $\det A^*=\overline{\det A}$.

0.3.2 交错和 受上述低维例子的启发，对于 $A=[a_{ij}]\in M_n(\mathbf{F})$，还有

$$\det A=\sum_{\sigma}\operatorname{sgn}\sigma\prod_{i=1}^{n}a_{i\sigma(i)},$$

其中，求和取遍 1，2，…，n 的所有 $n!$ 个排列 σ，排列 σ 的"正负号"或"正负号函数"$\operatorname{sgn}\sigma$ 是 +1 或 −1，取决于由(1，2，…，n)开始到得到排列 σ 所需对换(或两两交换)的最小数是偶数还是奇数．于是，每个乘积

$$a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}$$

都在行列式中出现，如果 σ 是偶排列，则在乘积前冠以+号，如果 σ 是奇排列，就冠以−号．

如果系数 $\operatorname{sgn}\sigma$ 用某些其他的函数来代替，那么所谓的广义矩阵函数就取代了 $\det A$．一个例子是 $\operatorname{per}A$，称为 A 的积和式(permanent)，其中 $\operatorname{sgn}\sigma$ 被恒等于 1 的函数所代替．

0.3.3 初等变换 有三种简单的基本变换，人们常常可以利用这些变换把任一个矩阵化简成与该矩阵相抵的、唯一简单的形式(即标准形)，以便用它来解线性方程组，计算行列式，矩阵求逆和研究矩阵的秩，等等．我们集中讨论关于行的变换，它们是如下三种变换．

第一种：交换矩阵的两行

交换第 i 行和第 j 行可以经左乘以矩阵

[8]

$$\begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & & & 1 & & \\ & & & & 1 & & & & \\ & & & & & \ddots & & & \\ & & & & & & 1 & & \\ & & & 1 & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ i\text{行} \\ \\ \\ \\ j\text{行} \\ \\ \\ \end{matrix}$$

i 列　　　j 列

来实现，其中，在 i，j 位置和 j，i 位置上的两个非对角元是 1，而所有未注明的元都是 0.

第二种：用一个非零纯量乘某一行

用一个纯量 c 乘 A 的第 i 行可以经左乘以矩阵

$$\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & c & & \\ & & & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ i\text{行} \\ \\ \\ \\ \end{matrix}$$

i 列

来实现，其中纯量 c 出现在 i，i 位置.

第三种：把某一行的纯量倍数加到另一行

[9] 用 c 乘第 i 行加到第 j 行相当于把 A 左乘以矩阵

$$\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & c & \ddots & \\ & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ j\text{行} \\ \\ \end{matrix},$$

i 列

其中纯量 c 出现在 j，i 位置. 注意，上述每个作初等变换的矩阵，正好是把相应的初等变换施于单位矩阵 I 的结果.

第一种初等变换在行列式上的作用是将行列式乘以 -1；第二种变换的作用是将它乘以纯量 c；第三种变换不改变行列式. 由此可知，如果一个矩阵有一个零行，或有两行相关，或任意 k 行相关，那么它的行列式为零. 一个矩阵的行列式是零，当且仅当它的诸行的一个子集线性相关.

0.3.4 行简化梯形阵 对每个 $A\in M_{m,n}(\mathbf{F})$，在 $M_{m,n}(\mathbf{F})$中存在一个标准形，A 的行简化梯形(RREF)，它可以经(不唯一的)一系列初等变换得到. 许多矩阵有相同的 RREF，每个矩阵不管经一列什么样的初等变换而得到，它只有一个 RREF. RREF 的定义是：

(a) 各非零行的第一个非零元是 1；

(b) 具有上述首元 1 的列的所有其他元都为零；

(c) 全由零元组成的行出现在矩阵的底部；

(d) 诸首元 1 位于从左到右的"阶梯型"之中，即下一行的首元 1 必须出现在其上一行的首元 1 的右边.

例如，

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & \pi \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

[10]

是 RREF. $A\in M_n(\mathbf{F})$的行列式是非零的，当且仅当它的 RREF 是单位矩阵

$$I = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix},$$

(它的行列式是 1). 记录下把 A 化成 RREF 的每个初等变换对 A 的行列式的作用，可以用来计算 $\det A$ 的值.

考虑线性方程组 $Ax=b$，其中 $A\in M_{m,n}(\mathbf{F})$和 $b\in\mathbf{F}^m$ 已知，而 $x\in\mathbf{F}^n$ 未知. 对 A 和 b 施以相同的初等变换，解集合不变. 解可以从$[A\ b]$的 RREF 看出来. 事实上，RREF 是唯一的，并且两个方程组 $Ax=b$ 是解等价的(有相同的解集合)，当且仅当两个增广矩阵$[A\ b]$有相同的 RREF.

后面，将讨论 RREF 在求矩阵的秩和逆中的作用.

0.3.5 乘法性质 行列式函数最关键和最重要的性质是，它是可相乘的：对于 A，$B\in M_n(\mathbf{F})$，

$$\det AB = \det A \det B.$$

用初等变换把 A 和 B 都行简化就可以证明这个等式.

0.3.6 行列式的函数特征 把行列式分别看成每一行(或列)的函数，而让其余各行(或列)不变，那么它是该行(或列)各元的线性函数. 从 Laplace 展开式来看，这是很明显的，因为一个给定的元的系数正好是它的上余子式，而余子式是固定不变的. 如果一个函数依次对其各变元的一个给定划分中的每一组变元都是线性的，就称这个函数是多重线性的. 这是相当广泛的一类函数. 例如，函数 $f(x_1, x_2)=x_1x_2$ 关于划分$\{x_1\}$，$\{x_2\}$是多重线性的. 又如，行列式作为矩阵的各元的函数，它对相应于各行(或列)的划分是多重线性的.

人们自然要问，是否迄今所提到的行列式的任何一组性质都把它描述成 $A\in M_n$ 的 n^2 个元

[11] 的一个纯量值函数. 行列式是唯一满足下列条件的函数 f: $M_n(\mathbf{F})\to\mathbf{F}$:

(a) 多重线性的;

(b) 交错的: 第一种初等变换使结果乘以 -1;

(c) 规范的: $f(I)=1$, 其中 $I\in M_n(\mathbf{F})$ 是单位矩阵.

积和式函数(这是另一类广义矩阵函数)也是多重线性的和规范的, 但它不是交错的.

0.4 秩

矩阵 $A\in M_{m,n}(\mathbf{F})$ 的秩是一个与它相关联的非负整数, 记作 rank A.

0.4.1 定义 如果 $A\in M_{m,n}(\mathbf{F})$, 那么, rank A 是 A 的列向量组中极大线性无关组的向量个数, 这个列向量组当然不是唯一的, 但这个集合的基数(元素个数)是唯一的. 值得注意的是, rank $A^T=$ rank A. 因此, 等价地, 秩可以用线性无关行来定义. 这通常简述为"行秩=列秩".

0.4.2 秩与线性方程组 线性方程组 $Ax=b$(0.3.4)可能有0个、1个或无限多个解, 但这些只是可能性. 如果方程组至少有一个解, 我们就称它是相容的. 线性方程组是相容的, 当且仅当 rank$[A b]=$rank A. $m\times(n+1)$ 矩阵 $[A b]$ 称为增广矩阵. 并且, 说增广矩阵与系数矩阵 A 有相同的秩就是说 b 是 A 的各列的线性组合. 这时, 把 b 添加到 A 的列中不会增加秩. 线性方程组 $Ax=b$ 的一个解是这样一个向量 x, 使得 b 是以其分量为系数的, A 的各列的线性组合.

0.4.3 RREF 和秩 初等变换不改变矩阵的秩. 因而, A 的秩与 A 的 RREF 的秩相同. 它恰好是 RREF 中非零行的个数. 用 RREF 计算秩受病态的影响: 在中间的数值计算中, 舍入误差可能使 RREF 的零行出现非零元, 因而影响秩的识别.

0.4.4 秩的特征 以下关于某个矩阵 $A\in M_{m,n}(\mathbf{F})$ 的各个命题都是相互等价的; 每个命题可能 [12] 在不同的上下文中用到.

(a) rank $A=k$;

(b) A 有 k 行组成的一个线性无关组, 而多于 k 行就线性相关;

(c) A 有 k 列组成的一个线性无关组, 而多于 k 列就线性相关;

(d) A 有一个其行列式不为零的 $k\times k$ 子矩阵, 但 A 的所有 $(k+1)\times(k+1)$ 子矩阵的行列式为0;

(e) A 的值域的维数是 k;

(f) 有 k 个但又不多于 k 个线性无关向量 b 组成的集合, 使线性方程组 $Ax=b$ 是相容的;

(g) $k=n$(A 的零空间的维数).

0.4.5 关于秩的不等式 涉及秩的不等式有以下几个.

(a) 对于 $A\in M_{m,n}(\mathbf{F})$, rank $A\leqslant\min\{m, n\}$.

(b) 从一个矩阵中划去若干行和(或)列后, 所得到子矩阵的秩不大于原矩阵的秩.

(c) 如果 $A\in M_{m,k}(\mathbf{F})$, $B\in M_{k,n}(\mathbf{F})$, 那么(rank $A+$rank B) $-k\leqslant$rank $AB\leqslant\min\{$rank A, rank $B\}$.

(d) 如果 $A, B\in M_{m,n}(\mathbf{F})$, 那么 rank$(A+B)\leqslant$rank $A+$rank B. 稍微复杂一点的是

Frobenius 的一个不等式.

(e) 如果 $A\in M_{m,k}(\mathbf{F})$，$B\in M_{k,p}(\mathbf{F})$，$C\in M_{p,n}(\mathbf{F})$，那么

$$\operatorname{rank} AB+\operatorname{rank} BC\leqslant \operatorname{rank} B+\operatorname{rank} ABC.$$

由此可推出另一些不等式.

0.4.6 关于秩的等式

(a) 如果 $A\in M_{m,n}(\mathbf{C})$，那么 $\operatorname{rank} A^*=\operatorname{rank} A^T=\operatorname{rank} \bar{A}=\operatorname{rank} A$.

(b) 如果 $A\in M_m(\mathbf{F})$和 $C\in M_n(\mathbf{F})$是非奇异的，且 $B\in M_{m,n}(\mathbf{F})$. 那么 $\operatorname{rank} AB=\operatorname{rank} B=\operatorname{rank} BC=\operatorname{rank} ABC$；即左乘或右乘以一个非奇异矩阵，其秩不变.

(c) 如果 A，$B\in M_{m,n}(\mathbf{F})$，那么，$\operatorname{rank} A=\operatorname{rank} B$，当且仅当存在非奇异矩阵 $X\in M_m(\mathbf{F})$和 $Y\in M_n(\mathbf{F})$，使得 $B=XAY$.

(d) 如果 $A\in M_{m,n}(\mathbf{C})$，那么 $\operatorname{rank} A^*A=\operatorname{rank} A$.

(e) 如果 $A\in M_{m,n}(\mathbf{F})$有秩 k，那么

$$A=XBY,$$

13

其中 $X\in M_{m,k}(\mathbf{F})$，$Y\in M_{k,n}(\mathbf{F})$和 $B\in M_k(\mathbf{F})$是非奇异的. 特别地，秩为 1 的矩阵 A 总可以写成形式 $A=xy^T$，其中 $x\in\mathbf{F}^m$，$y\in\mathbf{F}^n$ 是某两个向量.

0.5 非奇异性

如果一个线性变换或矩阵只对输入 0 才产生输出 0，就称它是非奇异的. 否则就称它是奇异的. 如果 $A\in M_{m,n}(\mathbf{F})$，且 $m<n$，则 A 一定是奇异的. 设 $A\in M_n(\mathbf{F})$，如果存在矩阵 $A^{-1}\in M_n(\mathbf{F})$(称为 A 的逆)使得 $A^{-1}A=I$，就称 A 是可逆的. 等价地，如果线性变换 A 是 1-1 的，且它的逆变换(它也是线性变换)存在，就称 A 是可逆的. 如果 $A\in M_n$，且 $A^{-1}A=I$，则 $AA^{-1}=I$；只要 A^{-1}存在，它就是唯一的.

可用多种不同的方法来判别 $A\in M_n(\mathbf{F})$是否非奇异，这是很有用的. 如果 $A\in M_n(\mathbf{F})$，以下各命题等价：

(a) A 是非奇异的；

(b) A^{-1}存在；

(c) $\operatorname{rank} A=n$；

(d) A 的各行线性无关；

(e) A 的各列线性无关；

(f) $\det A\neq 0$；

(g) A 的值域的维数为 n；

(h) A 的零空间的维数为 0；

(i) 对每个 $b\in\mathbf{F}^n$，$Ax=b$ 是相容的；

(j) 如果 $Ax=b$ 是相容的，那么解是唯一的；

(k) 对每个 $b\in\mathbf{F}^n$，$Ax=b$ 是唯一解；

(l) $Ax=0$ 的唯一解是 $x=0$；

(m) 0 不是 A 的特征值(见第 1 章).

$M_n(\mathbf{F})$中的非奇异矩阵构成一个群，称为一般线性群，常记作 $GL(n, \mathbf{F})$.

0.6 普通内积

我们约定，把 $\mathbf{F}^n$ 的元素看作列向量[即 $\mathbf{F}^n = M_{n,1}(\mathbf{F})$]. 于是，如果 $x \in \mathbf{C}^n$，则 x^T 和 x^* 是行向量. 注意，如果 $x \in \mathbf{R}^n$，则 $x^* = x^T$.

0.6.1 定义 纯量 y^*x 是 x 和 $y \in \mathbf{C}^n$ 的内积(纯量积)，常常把它记作 $\langle x, y\rangle \equiv y^*x$. 还可以定义不同于这个内积的各种内积，所以又把这个内积叫做向量空间 $\mathbf{C}^n$ 上的普通内积或标准内积. 应该指出 $\langle\cdot, \cdot\rangle$ 对第一个变元是线性的($\langle \alpha x_1 + \beta x_2, y\rangle = \alpha\langle x_1, y\rangle + \beta\langle x_2, y\rangle$ 对所有 α, $\beta \in \mathbf{C}$ 和 x_1, $x_2 \in \mathbf{C}^n$ 成立)，而对第二个变元是共轭线性的($\langle x, \alpha y_1 + \beta y_2\rangle = \bar{\alpha}\langle x, y_1\rangle + \bar{\beta}\langle x, y_2\rangle$ 对所有 α, $\beta \in \mathbf{C}$ 和 y_1, $y_2 \in \mathbf{C}^n$ 成立). [11]

0.6.2 正交性 两个向量 x, $y \in \mathbf{C}^n$ 称为正交，是指 $\langle y, x\rangle = 0$. 在二维和三维情形，通常把正交几何解释为垂直. 如果向量组 $\{x_1, x_2, \cdots, x_k\} \subset \mathbf{C}^n$ 中每两个向量都正交，就称它为正交组. 如果一个正交向量组中没有一个是零向量，那么它一定线性无关.

0.6.3 Cauchy-Schwarz 不等式 如果 $x \in \mathbf{C}^n$，非负纯量 $\langle x, x\rangle^{1/2}$ 是 x 的 Euclid 长度. Euclid 长度为 1 的向量叫做正规化向量(或单位向量). 对于任意一个非零向量 $x \in \mathbf{C}^n$，$x/\langle x, x\rangle^{1/2}$ 是与 x 同向的正规化向量. 基本的 Cauchy-Schwarz 不等式是说，

$$|\langle y, x\rangle| \leqslant \langle x, x\rangle^{1/2}\langle y, y\rangle^{1/2}$$

对所有 x, $y \in \mathbf{C}^n$ 成立；其中等号才成立，当且仅当 x 和 y 线性相关，推广正交概念，两个非零向量 x, $y \in \mathbf{C}^n$ 之间的夹角可明确地由公式

$$\cos\theta = \frac{|\langle y, x\rangle|}{\langle x, x\rangle^{1/2}\langle y, y\rangle^{1/2}}, \quad 0 \leqslant \theta \leqslant \frac{\pi}{2}$$

来定义.

0.6.4 Gram-Schmidt 标准正交化 一个线性无关向量组(它组成它的张成空间的一个基)可以用同一空间的一个标准正交(两两正交的，各自正规化的)基来代替，直观上看，这可能是合理的. 虽然这种替代原则上可以用无穷多种方式来实现，但是，在进行这种替代时，有一种非常简便的、行之有效的算法，称之为 Gram-Schmidt(标准正交化)过程，设 x_1, $\cdots$, x_n 是一个复向量空间的 n 个线性无关向量的集合. $\{z_1, \cdots, z_n\}$ 是要确定的正规化向量的正交组. 可以依次如下计算 z_i. 设 $y_1 = x_1$，并取

$$z_1 = \frac{y_1}{\langle y_1, y_1\rangle^{1/2}},$$

[15] 使得 z_1 是正规化的. 设 $y_2 = x_2 - \langle x_2, z_1\rangle z_1$，使得 y_2 与 z_1 正交，然后取

$$z_2 = \frac{y_2}{\langle y_2, y_2\rangle^{1/2}},$$

使得 z_2 是正规化的且与 z_1 正交. 类似地，将上述过程继续下去. 假设 z_1, $\cdots$, z_{k-1} 已被

确定，设

$$y_k = x_k - \langle x_k, z_{k-1}\rangle z_{k-1} - \langle x_k, z_{k-2}\rangle z_{k-2} - \cdots - \langle x_k, z_1\rangle z_1,$$

使得 y_k 与 z_1，…，z_{k-1} 正交. 然后正规化 y_k 便得到

$$z_k = \frac{y_k}{\langle y_k, y_k\rangle^{1/2}}.$$

继续做下去，直到得到所要求的标准正交向量 z_1，…，z_n. 注意，利用上述方法，一个无限的标准正交组能够由无限维向量空间中的一个无限可数线性无关组得到.

在 Gram-Schmidt 过程的每一步中，标准正交向量 z_1，…，z_k 仅仅是原无关向量 x_1，…，x_k 的线性组合(反之亦然). 如果记 $Z=[z_1 z_2 \cdots z_n]$ 和 $X=[x_1 x_2 \cdots x_n]$ 分别为以向量 z_i 和 x_i 为列的矩阵，那么 $Z=XR$，其中，矩阵 $R=[r_{ij}]$ 是非奇异上三角矩阵；即，当 $i>j$ 时，$r_{ij}=0$.

最后，我们指出，Gram-Schmidt 过程可以应用于任一有限的或可数的向量序列(不一定是线性无关的). 如果该集合不线性无关，那么，对于使 $\{x_1，\cdots，x_k\}$ 为线性相关组的最小 k 值，将得到向量 $y_k=0$. 这时，x_k 是 x_1，…，x_{k-1} 的线性组合. 用 x_{k+1} 代替 x_k，并继续 Gram-Schmidt 过程便可回答这样一个问题：什么是 $\{x_1，\cdots，x_n\}$ 的张成空间的基或维数?

0.6.5 标准正交基 一个标准正交向量组就是这样的正交向量组，它的每一个向量都是正规化的. 这个向量组不可能包含向量 0，并且一定线性无关. 一个标准正交基是由标准正交向量组成的基. 因为任一个基可(经 Gram-Schmidt)化成标准正交基，所以，任一有限维复向量空间有一个标准正交基. 因为在计算内积时，交叉项都变为零，所以与这样的基打交道是很合意的.

0.6.6 正交补 给定任意子集 $S\subset \mathbf{C}^n$，S 的正交补是集合

$$S^\perp \equiv \{x \in \mathbf{C}^n : x^* y = 0, \text{对所有 } y \in S\}.$$

[16]

即使 S 不是子空间，$S^\perp$ 也总是子空间. 我们有 $(S^\perp)^\perp = \mathrm{Span}\, S$，并且，当 S 是子空间时，$(S^\perp)^\perp = S$. 这时，$\dim S^\perp + \dim(S^\perp)^\perp = n$. 在关于线性方程组 $Ax=b$ 的叙述中(其中 $A\in M_{m,n}$)，应注意的是，A 的值域正好是 A^* 的零空间的正交补，即 $Ax=b$ 有解(不必唯一)，当且仅当 $b^* z=0$ 对所有使 $A^* z=0$ 的 $z\in \mathbf{C}^m$ 成立.

0.7 分块矩阵

类似于集合的划分，一个矩阵的分块是把该矩阵完全地分成一些互不相交的子矩阵，使得原矩阵的每一个元落到且只落到一个分块子矩阵中. 为识别一些有用的结构，矩阵分块往往是一个方便的方法.

0.7.1 子矩阵 设 $A\in M_{m,n}(\mathbf{F})$，对于指标集 $\alpha\subseteq\{1，\cdots，m\}$ 和 $\beta\subseteq\{1，\cdots，n\}$，把 A 中位于标号 α 的诸行和标号为 β 的诸列的(子)矩阵记作 $A(\alpha, \beta)$. 例如

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}(\{1,3\},\{1,2,3\}) = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}.$$

如果 $m=n$ 且 $\beta=\alpha$，子矩阵 $A(\alpha, \alpha)$ 称为 A 的主子矩阵，简记为 $A(\alpha)$. 说明一个子矩阵或一个主子矩阵是经划去那些行或列得来的，常常要比说它们包括那些行或列要方便些. 这可以通

过补充指标集来完成．例如，$A(\alpha',\beta')$是划去标号为α的诸行与标号为β的诸列后的结果．

矩阵A的子方阵的行列式称为A的一个子式．如果子矩阵是一个主子矩阵，那么其子式称为主子式．那些出现在 Laplace 展开式(0.3.1)中的带正负号的子式$[(-1)^{i+j}\det A_{ij}]$称为A的代数余子式．约定，空主子式是 1，即$\det A(\phi)=1$．

0.7.2 分块与乘法 如果$\alpha_1,\cdots,\alpha_t$组成$\{1,\cdots,m\}$的一个划分，而$\beta_1,\cdots,\beta_s$组成$\{1,\cdots,n\}$的一个划分，那么诸矩阵$A(\alpha_i,\beta_j)$，$1\leqslant i\leqslant t$，$1\leqslant j\leqslant s$，构成$A\in M_{m,n}(\mathbf{F})$的一个**分块**．如果$A\in M_{m,n}(\mathbf{F})$和$B\in M_{n,p}(\mathbf{F})$已经分块，且使$\{1,\cdots,n\}$的两个划分重合，那么就说这两个矩阵分块是**共形的**．这时，

17

$$[AB](\alpha_i,\gamma_j)=\sum_{k=1}^{s}A(\alpha_i,\beta_k)B(\beta_k,\gamma_j),$$

其中$A(\alpha_i,\beta_k)$和$B(\beta_k,\gamma_j)$是A和B的共形分块．等式左边是，乘积AB的一个子矩阵(按通常方式计算)，而右边的每一个被加项是一个标准的矩阵乘积．因此，共形的分块矩阵的乘法与通常的矩阵乘法相仿．当各被加项有相同的分块时．分块矩阵的加法也是有意义的．

0.7.3 分块矩阵的逆 求出非奇异分块矩阵A的逆的相应子块，即用相应的分块形式给出分块矩阵的逆，有时很有用．可以采用各种不同的，但彼此等价的方式来求分块矩阵的逆——假定$A\in M_n(\mathbf{F})$的某些子矩阵及A^{-1}也是非奇异的．为简单起见，设A是如下的分块

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},$$

其中，$A_{ii}\in M_{n_i}(\mathbf{F})$，$i=1,2$且$n_1+n_2=n$．关于$A^{-1}$的相应分块形式有一个有用的表示式

$$\begin{bmatrix}[A_{11}-A_{12}A_{22}^{-1}A_{21}]^{-1} & A_{11}^{-1}A_{12}[A_{21}A_{11}^{-1}A_{12}-A_{22}]^{-1}\\ [A_{21}A_{11}^{-1}A_{12}-A_{22}]^{-1}A_{21}A_{11}^{-1} & [A_{22}-A_{21}A_{11}^{-1}A_{12}]^{-1}\end{bmatrix},$$

其中，假定所有有关的逆存在．或者，用一般的指标集记号，可以记

$$A^{-1}(\alpha)=[A(\alpha)-A(\alpha,\alpha')A(\alpha')^{-1}A(\alpha',\alpha)]^{-1},$$

以及

$$A^{-1}(\alpha,\alpha')=A(\alpha)^{-1}A(\alpha,\alpha')[A(\alpha',\alpha)A(\alpha)^{-1}A(\alpha,\alpha')-A(\alpha')]^{-1},$$

仍假定有关的逆存在．还可以写出其余的表示式．注意，$A^{-1}(\alpha)$是A^{-1}的子矩阵，而$A(\alpha)^{-1}$是A的一个子矩阵的逆，并且这两个矩阵一般不相同．

0.7.4 小秩修正矩阵的逆 如果已知某个矩阵的逆，了解该矩阵再加上一个"小"秩矩阵时，其逆如何变化，这同样是一个有意义的问题．有这样的简便公式，只要修正矩阵的形式足够简单，就可以使新逆的计算比从头开始计算要简便．设非奇异矩阵$A\in M_n(\mathbf{F})$的逆A^{-1}已知，考虑

18

$$B=A+XRY,$$

其中，X是$n\times r$矩阵，而R是$r\times r$非奇异矩阵．如果B是非奇异的，那么

$$B^{-1}=A^{-1}-A^{-1}X(R^{-1}+YA^{-1}X)^{-1}YA^{-1}.$$

如果r比n小得多，那么求R和$R^{-1}+YA^{-1}X$的逆可能要比求B的逆更为容易，并且，如果

A 是容易求逆的，且有使乘法简化的形式．那么，采用这个公式可能要胜过直接求 B 的逆．例如，如果修正矩阵有秩 1，X 是 $n\times 1$ 的，Y 是 $1\times n$ 的，且 $R=[1]$，上述公式就变成

$$B^{-1}=A^{-1}-\frac{1}{1+YA^{-1}X}A^{-1}XYA^{-1},$$

(注意，此时 $XY=B-A$)．特别地，如果

$$B=I+xy^{T}$$

对于 x，$y\in\mathbf{F}^n$，$I\in M_n(\mathbf{F})$ 成立，那么，只要 $y^{T}x\neq -1$，就有

$$B^{-1}=I-\frac{1}{1+y^{T}x}xy^{T}.$$

0.8 行列式(续)

关于行列式的一些补充材料和恒等式对于某些论题的阐述很有用．其中大部分内容在基础线性代数中不易找到．

0.8.1 复合矩阵 设矩阵 $A\in M_{m,n}(\mathbf{F})$，其某个阶数的所有子式组成的阵列称为 A 的复合矩阵．特别地，它的 α，β 元为 $\det A(\alpha,\beta)$ 的 $\binom{m}{k}\times\binom{n}{k}$ 矩阵叫做 A 的第 k 次复合矩阵，记作 $C_k(A)$．这里，$\alpha\subseteq\{1,\cdots,m\}$ 和 $\beta\subseteq\{1,\cdots,n\}$ 都是基数为 $k\leqslant\min\{m,n\}$ 的指标集，按通常的辞典式次序排列．即 $\{1,2,4\}$ 排在 $\{1,2,5\}$ 之前，$\{1,2,5\}$ 排在 $\{1,3,4\}$ 之前，等等．例如，如果

$$A=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix},$$

则

$$C_2(A)=\begin{bmatrix}\det\begin{bmatrix}1&2\\4&5\end{bmatrix}&\det\begin{bmatrix}1&3\\4&6\end{bmatrix}&\det\begin{bmatrix}2&3\\5&6\end{bmatrix}\\ \det\begin{bmatrix}1&2\\7&8\end{bmatrix}&\det\begin{bmatrix}1&3\\7&9\end{bmatrix}&\det\begin{bmatrix}2&3\\8&9\end{bmatrix}\\ \det\begin{bmatrix}4&5\\7&8\end{bmatrix}&\det\begin{bmatrix}4&6\\7&9\end{bmatrix}&\det\begin{bmatrix}5&6\\8&9\end{bmatrix}\end{bmatrix}=\begin{bmatrix}-3&-6&-3\\-6&12&-6\\-3&-6&-3\end{bmatrix}.$$

如果 $A\in M_{m,k}(\mathbf{F})$，且 $B\in M_{k,n}(\mathbf{F})$，那么

$$C_r(AB)=C_r(A)C_r(B),\quad r\leqslant\min\{m,k,n\}.$$

另外，还有

$$C_r(tA)=t^rC_r(A),\quad t\in\mathbf{F}.$$

$$\text{若 } I\in M_n,\text{则 } C_r(I)=I\in M_{\binom{n}{r}}.$$

$$\text{若 } A\in M_n \text{ 是非奇异的，则 } C_k(A)^{-1}=C_k(A^{-1}).$$

$$\text{若 } A\in M_{m,n}(\mathbf{F})\text{，则 } C_k(A^T)=C_k(A)^T.$$

及

$$\text{若 } A\in M_{m,n}(\mathbf{C})\text{，则 } C_k(A^*)=C_k(A)^*.$$

0.8.2 经典伴随和逆 如果 $A\in M_n(\mathbf{F})$，由诸代数余子式

$$b_{ij}=(-1)^{i+j}\det A(\{j\}',\{i\}')$$

组成的转置矩阵 $B=[b_{ij}]\in M_n(\mathbf{F})$ 称为 A 的（经典）伴随，常记作 adj A. 有时用转置伴随来代替伴随，以免与 Hermite 伴随 A^* 相混淆. 注意，

$$\operatorname{adj} A = EC_{n-1}(A)^TE^{\ominus}.$$

其中

$$E=\begin{bmatrix}1 & & & & & 0\\ & -1 & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & \ddots & \\ 0 & & & & & +1\end{bmatrix}.$$

用 Laplace 展开式计算行列式的公式说明.

$$(\operatorname{adj} A)A = A(\operatorname{adj} A) = (\det A)I.$$

[20] 因而，如果 A 是非奇异的（$\det A\neq 0$），那么

$$A^{-1}=\frac{1}{\det A}\operatorname{adj} A.$$

一般说来，用伴随来数值计算矩阵的逆不是可取的方法. 但是伴随对于给出逆的解析表达式是有用的.

⊖ 原书给出的这个关系式是错误的（但书中尚未用到这个关系式）、adj A 与 $C_{n-1}(A)$ 应是如下关系：

$$\operatorname{adj} A = J_0C_{n-1}(A)^TJ_1 = J_1C_{n-1}(A)^TJ_0 \quad (n\text{ 为偶数})$$

$$\operatorname{adj} A = J_0C_{n-1}(A)^TJ_0 = J_1C_{n-1}(A)^TJ_1 \quad (n\text{ 为奇数}).$$

其中 n 阶方阵 J_0 和 J_1 分别为：

$$J_0=\begin{bmatrix} & & & & 1\\ 0 & & & -1 & \\ & & 1 & & \\ & -1 & & & \\ & \iddots & & 0 & \\ \pm 1 & & & & \end{bmatrix},\quad J_1=\begin{bmatrix} & & & & -1\\ & & & 1 & \\ & & -1 & & \\ & 1 & & & \\ \iddots & & & & \\ \mp 1 & & & & \end{bmatrix}.$$

详见引文：杜忠复等. 复合矩阵与伴随矩阵的关系及应用. 工科数学，1999，6，156-158. —译者注

0.8.3 Cramer 法则 当 $A\in M_n(\mathbf{F})$ 非奇异时，Cramer 法则是求线性方程组 $Ax=b$ 唯一解的一种方法. 它和逆的伴随表示有相同的计算手续，一般，只有在需要解析地给出解向量的个别解析分量时，这种方法才有用. 如果 x_i 是解向量 $x\in\mathbf{F}^n$ 的第 i 个分量，那么，Cramer 法则可述为公式

$$x_i=\frac{\det(A\underset{i}{\leftarrow}b)}{\det A}.$$

记号 $A\underset{i}{\leftarrow}b$ 表示 M_n 中第 i 列是 b，其余各列与 A 的对应列相同的矩阵. Cramer 法则可直接由行列式的乘法性质推出. 把方程组 $Ax=b$ 改写成

$$A(I\underset{i}{\leftarrow}x)=A\underset{i}{\leftarrow}b,$$

然后两边取行列式(利用乘法性质)可得

$$(\det A)\det(I\underset{i}{\leftarrow}x)=\det(A\underset{i}{\leftarrow}b).$$

而 $\det(I\underset{i}{\leftarrow}x)=x_i$，因而公式得证.

0.8.4 逆的子式 推广非奇异矩阵的逆的伴随公式，有如下重要公式：

$$\det A^{-1}(\alpha',\beta')=(-1)^{\left(\sum_{i\in\alpha}i+\sum_{j\in\beta}j\right)}\frac{\det A(\beta,\alpha)}{\det A},$$

它把 A^{-1} 的诸子式与 $A\in M_n(\mathbf{F})$ 的诸子式联系起来. 对于主子阵，这个公式有简单的形式

$$\det A^{-1}(\alpha')=\frac{\det A(\alpha)}{\det A}.$$

0.8.5 Schur 补和行列式公式 对于给定的矩阵 $A\in M_n(\mathbf{F})$，设 $\alpha\subseteq\{1,\cdots,n\}$ 是使 $A(\alpha)$ 非奇异的指标集. 记 $A(\alpha)$ 的逆为 $A(\alpha)^{-1}$. 用 α 和 α' 对 A 作 2×2 分块，据此，关于 $\det A$ 的重要公式是

21

$$\det A=\det A(\alpha)\det[A(\alpha')-A(\alpha',\alpha)A(\alpha)^{-1}A(\alpha,\alpha')].$$

注意，这个公式推广了(0.3.1)中的关于 2×2 矩阵行列式的普通公式. 称特殊矩阵

$$A(\alpha')-A(\alpha',\alpha)A(\alpha)^{-1}A(\alpha,\alpha')$$

为 A 的 Schur 补. 将

$$\begin{bmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{bmatrix}\text{右乘以}\begin{bmatrix}I&-A_{11}^{-1}A_{12}\\0&I\end{bmatrix},$$

然后把 A_{11} 和 $A(\alpha)$ 等同起来，就可验证 $\det A$ 和 Schur 补公式成立. 注意，Schur 补已在 A^{-1} 的分块中出现过[见(0.7.3)].

0.8.6 Sylvester 恒等式 设 $\alpha\subseteq\{1,\cdots,n\}$ 是固定的指标集，设 $B=[b_{ij}]\in M_{n-k}(\mathbf{F})$ 由

$$b_{ij}=\det A(\alpha\cup\{i\},\alpha\cup\{j\})$$

所定义，其中 k 是 α 的基数，$i,j\in\{1,\cdots,n\}$ 是不包含在 α 中的指标，$A\in M_n(\mathbf{F})$. 另一个有用的行列式恒等式是

$$\det B=[\det A(\alpha)]^{n-k-1}\det A$$

0.8.7 Cauchy-Binet 公式 这个有用的公式是可以想起来的，这是因为它看上去与矩阵的乘法公式相类似. 它等价于复合矩阵的乘法性质(见 0.8.1)，所以这不是偶然的巧合. 设 $A\in M_{m,k}(\mathbf{F})$，$B\in M_{k,n}(\mathbf{F})$ 和 $C=AB$. 再设 $1\leqslant r\leqslant\min\{m,k,n\}$，$\alpha\subseteq\{1,\cdots,m\}$ 和 $\beta\subseteq\{1,\cdots,n\}$ 都是基

数为 r 的指标集. 关于 C 的 α，β 子式的表示式是

$$\det C(\alpha,\beta)=\sum_{\gamma}\det A(\alpha,\gamma)\det B(\gamma,\beta),$$

其中和式取遍基数为 r 的所有指标集 $\gamma\subseteq\{1,\cdots,k\}$.

0.8.8 子式间的关系 已知 $A\in M_{m,n}(\mathbf{F})$，给定基数为 k 的固定指标集 $\alpha\subseteq\{1,\cdots,m\}$，当 $\omega\subseteq\{1,\cdots,n\}$ 取遍基数为 k 的有序指标集时，诸子式

$$\det A(\alpha,\omega)$$

不是代数无关的，因为在各子矩阵中诸子式多于诸子矩阵中的各不相同的元. 在这些子式当中，二次关系是知道的. 设 $i_1,i_2,\cdots,i_k\in\{1,\cdots,n\}$ 是 k 个互异的指标，不一定取自然顺序，又设

22

$$A(\alpha;i_1,\cdots,i_k)$$

表示这样一个矩阵，它的各行用 α 标号，而它的第 j 列是 $A(\alpha,\{1,\cdots,n\})$ 的 i_j 列. 这与前述记号的差别是，列可以不按自然顺序，如在 $A(\{1,3\};4,2)$ 中，它的第1列有 A 中的1，4元和3，4元. 于是有关系式

$$\det A(\alpha;i_1,\cdots,i_k)\det A(\alpha;j_1,\cdots,j_k)$$

$$=\sum_{t=1}^{k}\det A(\alpha;i_1,\cdots,i_{s-1},j_t,i_{s+1},\cdots,i_k)\det A(\alpha;j_1,\cdots,j_{t-1},i_s,j_{t+1},\cdots,j_k),$$

它对于每个 $s=1,\cdots,k$ 和互异指标的所有序列

$$i_1,\cdots,i_k\in\{1,\cdots,n\}\text{ 和 }j_1,\cdots,j_k\in\{1,\cdots,n\}$$

成立.

0.9 矩阵的特殊形式

经常会遇到某些特殊形式的矩阵，这些矩阵具有特殊的性质. 为了参阅，其中有些矩阵值得在这里介绍，并给出其名称.

0.9.1 对角矩阵 如果矩阵 $D=[d_{ij}]\in M_n$ 在 $j\neq i$ 时，有 $d_{ij}=0$，就称 D 为对角矩阵. 平常，把这个矩阵记作 $D=\operatorname{diag}(d_{11},\cdots,d_{nn})$ 或 $D=\operatorname{diag} d$，其中 d 是 D 的对角元组成的向量. 如果一个对角矩阵的所有对角元都是正(非负)实数，就称它为正(非负)对角矩阵. 注意，术语正对角矩阵指的是，矩阵是对角形的，且有正对角元；它不是指那种所有对角元碰巧都是正数的一般矩阵. 单位矩阵是正对角矩阵的一个例子. 如果对角矩阵 D 的诸对角元都相等，就称 D 为纯量矩阵；即对某 $\alpha\in\mathbf{C}$ 有 $D=\alpha I$. 一个矩阵左乘或右乘以一个纯量矩阵，与用相应的纯量乘这个矩阵的作用相同.

一个对角矩阵的行列式正好是它的诸对角元之乘积：$\det D=\prod_{i=1}^{n}d_{ii}$. 因而，一个对角矩阵是非奇异的，当且仅当它的所有对角元是非零的. $A\in M_n$ 左乘以对角阵 D，即 DA，就是用 D 的诸对角元乘 A 的各行(A 的第 i 行乘以 d_{ii}，$i=1,2,\cdots,n$). 而右乘以 D，即 AD，就是

用 D 的诸对角元乘 A 的各列. 因此，所有对角矩阵关于乘法相互交换，并且，一个对角矩阵与某个矩阵 $A=[a_{ij}]\in M_n$ 可交换，当且仅当在 D 的第 i 个对角元与第 j 个对角元不相同时，有 $a_{ij}=0$. 两个对角矩阵的乘积也是对角矩阵，其对角元正好是它们相应的对角元的两两乘积. 类似地，可以规定一个对角矩阵的正整数幂. [23]

0.9.2 分块对角矩阵 具有形式

$$A=\begin{bmatrix} A_{11} & & 0 \\ & A_{22} & \\ & & \ddots \\ 0 & & & A_{kk} \end{bmatrix}$$

的矩阵 $A\in M_n$ 称为分块对角矩阵，其中，$A_{ii}\in M_{n_i}$，$i=1,2,\cdots,k$ 且 $\sum_{i=1}^{k} n_i=n$. 形式上，这个矩阵常常用 $A=A_{11}\oplus A_{22}\oplus\cdots\oplus A_{kk}$ 来表示，或简记作 $\oplus\sum_{i=1}^{k} A_{ii}$，称这个矩阵为 A_{11}，A_{22}，…，A_{kk} 的直和. 从分块矩阵的乘法来考虑，分块对角矩阵的许多性质推广了对角矩阵的性质，例如，$\det(\oplus\sum_{i=1}^{k} A_{ii})=\prod_{i=1}^{k}\det A_{ii}$，因而，$A=\oplus\sum A_{ii}$ 是非奇异的，当且仅当每个 A_{ii} 是非奇异的，$i=1,2,\cdots,k$. 另外，直和 $A=\oplus\sum_{i=1}^{k} A_{ii}$ 与 $B=\oplus\sum_{i=1}^{k} B_{ii}$ 可交换，其中 A_{ii}，B_{ii} 是同阶的. 当且仅当 A_{ii} 与 B_{ii} 可交换，$i=1,2,\cdots,k$. 还有，$\operatorname{rank}(\oplus\sum_{i=1}^{k} A_{ii})=\sum_{i=1}^{k}\operatorname{rank} A_{ii}$.

0.9.3 三角矩阵 如果矩阵 $T=[t_{ij}]\in M_n$，当 $j<i$ 时，有 $t_{ij}=0$，就称 T 为上三角矩阵. 如果当 $j\leqslant i$ 时，$t_{ij}=0$，就称 T 是严格上角矩阵，类似地，T 称为下三角(或严格下三角)矩阵，是指它的转置是上三角(或严格上三角)矩阵. 与对角矩阵类似，三角矩阵的行列式是它的诸对角元之乘积. 三角矩阵(两种中的任一种)不一定与另一种三角矩阵可交换. $A\in M_n$ 左乘以下三角矩阵 L，即 LA，就是用 L 的第 1 行至第 i 行的线性组合代替 A 的第 i 行. 在述及三角矩阵时，有时采用术语右(代替上)和左(代替下)三角矩阵. 三解矩阵的秩至少是(也可能大于)主对角线上非零元的个数. [24]

0.9.4 分块三角矩阵 具有形状

$$A=\begin{bmatrix} A_{11} & & * \\ & A_{22} & \\ 0 & \ddots & \\ & & A_{kk} \end{bmatrix}$$

的矩阵 $A\in M_n$ 称为分块上三角矩阵，其中 $A_{ii}\in M_{n_i}$，$i=1,\cdots,k$，$\sum_{i=1}^{k} n_i=n$，而“$*$”表示任意块元. 分块下三角矩阵、严格分块下三角矩阵和严格分块上三角矩阵都可以类似地定义. 分块三角矩阵的行列式是诸对角子块的行列式之积. 分块三角矩阵的秩至少是(也可能大于)诸对角子块的秩之和.

0.9.5 置换矩阵 如果矩阵 $P\in M_n$ 在它的每一行和每一列正好有一个元等于 1，而其余所有的元都是 0，就称 P 为置换矩阵. 乘以这样一个矩阵的效果是使被乘矩阵的行或列互换. 例如

$$P=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\in M_3$$

是置换矩阵，而

$$P\begin{bmatrix}1\\2\\3\end{bmatrix}=\begin{bmatrix}2\\1\\3\end{bmatrix}$$

是向量 $\begin{bmatrix}1\\2\\3\end{bmatrix}$ 的行(分量)的互换，即它把第一项换到第 2 个位置，把第 2 项换到第 1 个位置，而让第 3 项保持在第 3 个位置. 一般地，矩阵 $A\in M_{m,n}$ 左乘以置换矩阵 $P\in M_m$ 就是互换 A 的行，而矩阵 $A\in M_{m,n}$ 右乘以置换矩阵 $P\in M_n$ 就是互换 A 的列. 实施(0.3.3)的第一种初等变换的矩阵是一个特殊形式的置换矩阵，称之为对换矩阵.

置换矩阵的行列式是 ± 1 [在(0.3.2)的公式中正好有一个被加项非零]，因而置换矩阵一定是非奇异的. 虽然置换矩阵关于乘法一般是不交换的，但两个置换矩阵的乘积还是一个置换矩阵. 因为单位矩阵是一个置换矩阵，并且对每个置换 P，有 $P^T=P^{-1}$. 所以，置换矩阵构成 M_n 中的非奇异矩阵群 $GL(n, \mathbf{C})$ 的一个子群，它具有有限基数 $n!$. 事实上，任一置换矩阵是一些对换矩阵的乘积.

25

因为，如果置换矩阵 $P\in M_n$ 以某种方式互换行，则 $P^T=P^{-1}$ 就以同一方式互换列，所以，变换 $A\to PAP^T$ 以相同的方式互换 $A\in M_n$ 的行和列，这个变换相当于重排诸元的足码. 如果矩阵 $A\in M_n$ 有某个置换矩阵 P 使得 PAP^T 是三角矩阵，就称 A 为本性三角矩阵. 这些矩阵与三角矩阵有许多共同之处.

0.9.6 轮换矩阵 具有形状

$$A=\begin{bmatrix}a_1&a_2&&\cdots&a_n\\a_n&a_1&a_2&\cdots&a_{n-1}\\a_{n-1}&a_n&a_1&\cdots&a_{n-2}\\\vdots&\vdots&\ddots&\ddots&\vdots\\a_2&a_3&\cdots&a_n&a_1\end{bmatrix}$$

的矩阵称为轮换矩阵. 每一行正好是前一行循环一步，使得每一行各元刚好是前一行的各元的一个循环排列. 置换矩阵

$$C=\begin{bmatrix}0&1&0&\cdots&0\\&0&1&&\vdots\\\vdots&&&&\\&&&\ddots&0\\&&&\ddots&\\0&&&&1\\1&0&&\cdots&0\end{bmatrix}$$

称为基本轮换转置矩阵. 矩阵 $A\in M_n$ 可以写成形式

$$A=\sum_{k=0}^{n-1}a_{k+1}C^k$$

当且仅当 A 是轮换矩阵. 这里 $C^0=I=C^n$，而系数 a_1，a_2，$\cdots$，a_n 正好是 A 的第 1 行的诸元. 因为这个表达式，轮换矩阵具有与 C 相关联的优美结构. 又因为 $C^n=I$，所以两个轮换矩阵的乘积还是一个轮换矩阵. 另外，轮换矩阵在乘法下交换. 轮换矩阵有几种推广，例如其中之一是，把各行向前(或向右)循环一个大于 1 的固定步数. [26]

0.9.7 Toeplitz 矩阵 具有形状

$$A=\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_n \\ a_{-1} & a_0 & a_1 & \cdots & a_{n-1} \\ a_{-2} & a_{-1} & a_0 & \cdots & a_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{-n+1} & a_{-n+2} & \ddots & a_0 & a_1 \\ a_{-n} & a_{-n+1} & \cdots & a_{-1} & a_0 \end{bmatrix}$$

的矩阵 $A=[a_{ij}]\in M_{n+1}$ 称为 Toeplitz 矩阵. 对于某个给定的序列 a_{-n}，a_{-n+1}，$\cdots$，a_{-1}，a_0，a_1，a_2，$\cdots$，a_{n-1}，$a_n\in\mathbf{C}$，一般项 $a_{ij}=a_{j-i}$. 沿着平行于主对角线的诸对角线从上到下，A 的各元取常值. Toeplitz 矩阵

$$B=\begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix} \quad 和 \quad F=\begin{bmatrix} 0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & \ddots & \\ 0 & & 1 & 0 \end{bmatrix}$$

称为“后向位移”矩阵和“前向位移”矩阵，这是因它们在标准基 $\{e_1, e_2, \cdots, e_{n+1}\}$ 的诸元上的作用而得名. 矩阵 $A\in M_{n+1}$ 可以写成形式

$$A=\sum_{k=1}^{n}a_{-k}F^k+\sum_{k=0}^{n}a_kB^k$$

当且仅当 A 是 Toeplitz 矩阵. 在涉及三角矩的问题中，自然要遇到 Toplitz 矩阵.

0.9.8 Hankel 矩阵 具有形状

$$A=\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & a_3 & \cdots & a_{n+1} \\ a_2 & a_3 & a_4 & \cdots & a_{n+2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n-1} & a_n & a_{n+1} & \cdots & a_{2n-1} \\ a_n & a_{n+1} & a_{n+2} & \cdots & a_{2n} \end{bmatrix}$$

[27]

的矩阵 $A\in M_{n+1}$ 称为 Hankel 矩阵. 对于某个给定的序列 a_0，a_1，a_2，$\cdots$，a_{2n-1}，a_{2n}，一般项

$a_{ij}=a_{i+1,j-1}$. 沿着与主对角线垂直的诸对角线，A 的各元取常值. 在涉及幂矩的问题中，自然要遇到 Hankel 矩阵. 注意，如果

$$P=\begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \cdot^{\cdot^{\cdot}} & & \\ 1 & & & \\ 1 & & & 0 \end{bmatrix}$$

称之为"后向单位"(置换)矩阵. 那么，对任意 Toeplitz 矩阵 T，PT 是 Hankel 矩阵，而对于任意 Hankel 矩阵 H，PH 是 Toeplitz 矩阵. 因为，$P=P^T=P^{-1}$ 和 Hankel 矩阵是对称的，这表明任一 Toeplitz 矩阵是两个对称矩阵(P 和 Hankel 矩阵)的乘积.

0.9.9 Hessenberg 矩阵 如果矩阵 $A=[a_{ij}]\in M_n$ 对于 $i>j+1$ 有 $a_{ij}=0$：

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & & a_{1n} \\ a_{21} & a_{22} & & & \\ 0 & a_{32} & \ddots & & \vdots \\ \vdots & 0 & \ddots & & \\ & \vdots & \ddots & & \\ 0 & 0 & \cdots 0 & a_{n,n-1} & a_{n,n} \end{bmatrix}$$

就称 A 呈上 Hessenberg 形状，或称 A 是上 Hessenberg 矩阵. 如果 A^T 是上 Hessenberg 矩阵，就称 $A\in M_n$ 为下 Hessenberg 矩阵.

0.9.10 三对角矩阵 如果矩阵 $A=[a_{ij}]\in M_n$ 既是上，又是下 Hessenberg 矩阵，就称之为三对角矩阵. 即，A 是三对角矩阵，是指当 $|i-j|>1$ 时，$a_{ij}=0$：

$$A=\begin{bmatrix} a_{11} & a_{12} & & & 0 \\ a_{21} & a_{22} & a_{23} & & \\ & a_{32} & a_{33} & & \\ & & & \ddots & \\ & & \ddots & \ddots & \\ & & & \ddots & a_{n-1,n} \\ 0 & & & a_{n,n-1} & a_{n,n} \end{bmatrix}.$$

[28] 用归纳法容易计算三对角矩阵的行列式. 注意

$$\begin{aligned} &\det A(\{1,2,\cdots,k+1\}) \\ &= a_{k+1,k+1}\det A(\{1,\cdots,k\}) - a_{k+1,k}a_{k,k+1}\det A(\{1,\cdots,k-1\}) \\ &\qquad\qquad k=2,\ \cdots,\ n-1. \end{aligned}$$

0.9.11 矩阵和 Lagrange 插值法 Vandermonde 矩阵 $A\in M_n(\mathbf{F})$ 是具有形状

$$A=\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^{n-1} \end{bmatrix} \tag{0.9.11.1}$$

的矩阵，其中 x_1，x_2，…，$x_n\in\mathbf{F}$；即 $A=[a_{ij}]$，其中 $a_{ij}=x_i^{j-1}$. 下述等式成立

$$\det A = \prod_{\substack{i,j=1\\ i>j}}^{n}(x_i - x_j), \qquad (0.9.11.2)$$

于是，一个 Vandermonde 矩阵是非奇异的，当且仅当 n 个参数 x_1，x_2，…，x_n 是互异的.

Vandermonde 矩阵出现在插值问题中，所谓插值问题，就是求系数在 $\mathbf{F}$ 中的、次数至多为 $n-1$ 的多项式 $p(x)=a_{n-1}x^{n-1}+a_{n-2}x^{n-2}+\cdots+a_1x+a_0$，使得

$$\begin{aligned}
p(x_1) &= a_0+a_1x_1+a_2x_1^2+\cdots+a_{n-1}x_1^{n-1}=y_1,\\
p(x_2) &= a_0+a_1x_2+a_2x_2^2+\cdots+a_{n-1}x_2^{n-1}=y_2,\\
&\ \vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots\\
p(x_n) &= a_0+a_1x_n+a_2x_n^2+\cdots+a_{n-1}x_n^{n-1}=y_n,
\end{aligned} \qquad (0.9.11.3)$$

其中 x_1，x_2，…，x_n 和 y_1，y_2，…，y_n 是中 $\mathbf{F}$ 的已知元素. 插值条件(0.9.11.3)是关于 n 个未知系数 a_0，a_1，…，a_{n-1} 的 n 个方程的组，并且它们有形式 $Aa=y$，其中 $a=[a_0, a_1, \cdots, a_{n-1}]^T\in\mathbf{F}^n$，$y=[y_1, y_2, \cdots, y_n]^T\in\mathbf{F}^n$，且 $A\in M_n(\mathbf{F})$ 是 Vandermonde 矩阵(0.9.11.1). 如果点 x_1，x_2，…，x_n 各不相同，这个插值问题总有一个解，因为这时 A 是非奇异矩阵.

如果点 x_1，x_2，…，x_n 是互异的，插值多项式的系数原则上可以通过解方程组(0.9.11.3)求得. 但是，用特殊的 Lagrange 插值多项式

[29]

$$L_i(x)=\frac{\prod\limits_{\substack{j=1\\ j\neq i}}^{n}(x-x_j)}{\prod\limits_{\substack{j=1\\ j\neq i}}^{n}(x_i-x_j)}, \quad i=1,2,\cdots,n$$

表示插值多项式 $P(x)$ 通常更为有效. 每个多项式 $L_i(x)$ 为 $n-1$ 次，且具有性质：如果 $k\neq i$，$L_i(x_k)=0$；而 $L_i(x_i)=1$. 因此，对于满足方程组(0.9.11.3)的、次数不超过 $n-1$ 的多项式 $p(x)$，有 Lagrange 插值公式

$$p(x) = y_1L_1(x)+y_2L_2(x)+\cdots+y_nL_n(x). \qquad (0.9.11.4)$$

0.10 基的变换

设 V 是域 $\mathbf{F}$ 上的 n 维向量空间，$\mathscr{B}_1=\{v_1, v_2, \cdots, v_n\}$ 是 V 的一个基. 如果 $x\in V$ 是任一给定的向量，因为集合 $\mathscr{B}_1$ 张成 V，则 x 有某个表示式 $x=\alpha_1v_1+\alpha_2v_2+\cdots+\alpha_nv_n$. 如果在同一个基下存在另一个表达式 $x=\beta_1v_1+\beta_2v_2+\cdots+\beta_nv_n$，那么，

$$0 = x-x = (\alpha_1-\beta_1)v_1+(\alpha_2-\beta_2)v_2+\cdots+(\alpha_n-\beta_n)v_n$$

由基 $\mathscr{B}_1$ 无关可知所有 $\alpha_i-\beta_i=0$. 给定基 $\mathscr{B}_1$，从 V 到 $\mathbf{F}^n$ 的线性映射

$$x\to[x]_{\mathscr{B}_1}=\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_n\end{bmatrix}, \quad 其中\ x=\alpha_1v_1+\alpha_2v_2+\cdots+\alpha_nv_n$$

是意义明确的、一对一的和到上的. 纯量 α_i 称为 x 关于基 $\mathscr{B}_1$ 的坐标，而列向量 $[x]_{\mathscr{B}_1}$ 是 x 的唯一 $\mathscr{B}_1$ 坐标表示.

设 $T: V\to V$ 是给定的线性变换. 只要知道 n 个向量 Tv_1，Tv_2，…，Tv_n，T 对任一 $x\in V$ 的作用就被确定了；这是因为任一 $x\in V$ 有唯一表示 $x=\alpha_1v_1+\cdots+\alpha_nv_n$，且由线性性质可知，$Tx=T(\alpha_1v_1+\cdots+\alpha_nv_n)=T(\alpha_1v_1)+\cdots+T(\alpha_nv_n)=\alpha_1Tv_1+\cdots+\alpha_nTv_n$. 因此，一旦知道 $[x]_{\mathscr{B}_1}$，就可以确定 Tx 的值.

设 $\mathscr{B}_2=\{w_1, w_2, \cdots, w_n\}$ 是 V 的另一个基，可能与 $\mathscr{B}_1$ 不同，并且假定 Tv_j 的 $\mathscr{B}_2$ 坐标表示是

$$[Tv_j]_{\mathscr{B}_2}=\begin{bmatrix}t_{1j}\\t_{2j}\\\vdots\\t_{nj}\end{bmatrix},\quad j=1,2,\cdots,n.$$

[30]

那么，对任一 $x\in V$，有

$$[Tx]_{\mathscr{B}_2}=\Big[\sum_{j=1}^{n}\alpha_jTv_j\Big]_{\mathscr{B}_2}=\sum_{j=1}^{n}\alpha_j[Tv_j]_{\mathscr{B}_2}$$

$$=\sum_{j=1}^{n}\alpha_j\begin{bmatrix}t_{1j}\\t_{2j}\\\vdots\\t_{nj}\end{bmatrix}=\begin{bmatrix}t_{11}&\cdots&t_{1n}\\\vdots&\ddots&\vdots\\t_{n1}&\cdots&t_{nn}\end{bmatrix}\begin{bmatrix}\alpha_1\\\vdots\\\alpha_n\end{bmatrix}.$$

$n\times n$ 阵列 $[t_{ij}]$ 依赖于 T 和基 $\mathscr{B}_1$ 和 $\mathscr{B}_2$ 的选择，但不取决于 x. 我们定义 T 的 $\mathscr{B}_1$–$\mathscr{B}_2$ 基表示为

$${}_{\mathscr{B}_2}[T]_{\mathscr{B}_1}=\begin{bmatrix}t_{11}&\cdots&t_{1n}\\\vdots&\ddots&\vdots\\t_{n1}&\cdots&t_{nn}\end{bmatrix}=[[Tv_1]_{\mathscr{B}_2}\cdots[Tv_n]_{\mathscr{B}_2}].$$

已经证明，对任一 $x\in V$，$[Tx]_{\mathscr{B}_2}={}_{\mathscr{B}_2}[T]_{\mathscr{B}_1}[x]_{\mathscr{B}_1}$. 实际上，为了给出 T 的一个基表示，$\mathscr{B}_2=\mathscr{B}_1$ 的情形是最常见的；${}_{\mathscr{B}_1}[T]_{\mathscr{B}_1}$ 称为 T 的 $\mathscr{B}_1$ 基表示.

考虑单位线性变换 $I: V\to V$，它定义为对所有 x，$Ix=x$. 于是，对所有 $x\in V$，有

$$[x]_{\mathscr{B}_2}=[Ix]_{\mathscr{B}_2}={}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}[x]_{\mathscr{B}_1}={}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}[Ix]_{\mathscr{B}_1}={}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}\,{}_{\mathscr{B}_1}[I]_{\mathscr{B}_2}[x]_{\mathscr{B}_2}.$$

依次取 $x=w_1$，w_2，…，w_n，这个恒等式能计算出 ${}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}\,{}_{\mathscr{B}_1}[I]_{\mathscr{B}_2}$ 的每一列，因而证明了

$${}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}\,{}_{\mathscr{B}_1}[I]_{\mathscr{B}_2}=\begin{bmatrix}1&&0\\&\ddots&\\0&&1\end{bmatrix}=I.$$

我们泛泛采用同一个记号 I 表示 $n\times n$ 单位矩阵和单位线性变换. 如果从 $[x]_{\mathscr{B}_1}=[Ix]_{\mathscr{B}_1}=\cdots$ 开始作同样的计算，也会得出

$${}_{\mathscr{B}_1}[I]_{\mathscr{B}_2}\,{}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}=I$$

因此，矩阵 ${}_{\mathscr{B}_1}[I]_{\mathscr{B}_2}$ 是矩阵 ${}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}$ 的逆矩阵. 如果记 $S={}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}$，那么 $S^{-1}={}_{\mathscr{B}_1}[I]_{\mathscr{B}_2}$. 于是，形如 ${}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}$ 的每一个矩阵都是可逆的. 反过来每一个可逆矩阵 $S=[s_1\, s_2 \cdots s_n]\in M_n(\mathbf{F})$ 对某个基 $\mathscr{B}$ 有形式 ${}_{\mathscr{B}_1}[I]_{\mathscr{B}}$. 可以把 $\mathscr{B}$ 看成是用 $[\tilde{s}_i]_{\mathscr{B}_1}=s_i$，$i=1$，2，…，$n$ 定义的向量组 $\{\tilde{s}_1, \tilde{s}_2, \cdots, \tilde{s}_n\}$. 因为 S 可逆，所以向量组 $\mathscr{B}$ 无关.

注意到 [31]

$${}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}=[[Iv_1]_{\mathscr{B}_2}\cdots[Iv_n]_{\mathscr{B}_2}]=[[v_1]_{\mathscr{B}_2}\cdots[v_n]_{\mathscr{B}_2}],$$

于是，${}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}$ 用基 $\mathscr{B}_2$ 表示基 $\mathscr{B}_1$ 的各个元素. 现在 $x\in V$，经计算

$$\begin{aligned}{}_{\mathscr{B}_2}[T]_{\mathscr{B}_2}[x]_{\mathscr{B}_2}&=[Tx]_{\mathscr{B}_2}=[I(Tx)]_{\mathscr{B}_2}={}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}[Tx]_{\mathscr{B}_1}\\&={}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}\,{}_{\mathscr{B}_1}[T]_{\mathscr{B}_1}[x]_{\mathscr{B}_1}={}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}\,{}_{\mathscr{B}_1}[T]_{\mathscr{B}_1}[Ix]_{\mathscr{B}_1}\\&={}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}\,{}_{\mathscr{B}_1}[T]_{\mathscr{B}_1}\,{}_{\mathscr{B}_1}[I]_{\mathscr{B}_2}[x]_{\mathscr{B}_2}.\end{aligned}$$

依次取 $x=w_1$，w_2，…，w_n，便得出

$${}_{\mathscr{B}_2}[T]_{\mathscr{B}_2}={}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}\,{}_{\mathscr{B}_1}[T]_{\mathscr{B}_1}\,{}_{\mathscr{B}_1}[I]_{\mathscr{B}_2}.$$

这个恒等式说明，如果用来计算表示的基发生变化，T 的基表示将如何变化. 由于这个缘故，才称矩阵 ${}_{\mathscr{B}_2}[I]_{\mathscr{B}_1}$ 为基 $\mathscr{B}_1$-$\mathscr{B}_2$ 的变换矩阵.

任一矩阵 $A\in M_n(\mathbf{F})$ 是某个线性变换 T：$V\to V$ 的一个基表示，这是因为，如果 $\mathscr{B}$ 是 V 的任一基，可以用 $[Tx]_{\mathscr{B}}=A[x]_{\mathscr{B}}$ 来确定 Tx. 不难算出，对这个 T，${}_{\mathscr{B}}[T]_{\mathscr{B}}=A$. [32]

第1章　特征值、特征向量和相似性

1.0　导引

在本章以及后面各章中，先引入该章要讨论的主要问题，并用例子说明它们是如何从理论上或应用中产生的.

1.0.1 基变换和相似性　每个可逆矩阵是基变换矩阵，而每个基变换矩阵是可逆矩阵[见(0.10)节]. 因此，如果 $\mathscr{B}$ 是向量空间 V 的一个给定的基，T 是 V 的一个给定的线性变换，且 $A={}_{\mathscr{B}}[T]_{\mathscr{B}}$ 是 T 的 $\mathscr{B}$ 基表示，那么，T 的所有可能基表示的集合是

$$\{{}_{\mathscr{B}_1}[I]_{\mathscr{B}}\,{}_{\mathscr{B}}[T]_{\mathscr{B}}\,{}_{\mathscr{B}}[I]_{\mathscr{B}_1} : \mathscr{B}_1 \text{ 是 } V \text{ 的基}\}$$

$$=\{S^{-1}AS : S\in M_n(\mathbf{F}) \text{ 是可逆矩阵}\}.$$

这正是与给定的矩阵 A 相似的所有矩阵的集合. 因此，相似的，而不是恒等的诸矩阵正好是同一个线性变换的不同的基表示.

人们自然希望相似的矩阵会有许多共同的性质——至少是基于线性变换的那些固有性质——这是线性代数的重要论题. 一个矩阵只是某个线性变换的所有可能表示中的一个. 从关于一个矩阵的问题去探讨关于该线性变换某些固有性质的问题，这样处理问题往往很见效.

33

相似概念是本章的主要概念.

1.0.2 约束极值和特征值　本章第二个主要概念是特征向量和特征值的概念. 我们将看到，使 Ax 是 x 的一个倍数的非零向量 x 在研究一般矩阵或线性变换的结构中起着重要的作用，而这样的向量出现在求具有一个几何约束条件的实对称二次型的极大值(或极小值)的基本问题中，即：

$$\text{假定 } x\in\mathbf{R}^n, x^Tx=1, \text{求 } x^TAx \text{ 的极大值},$$

其中 $A^T=A\in M_n(\mathbf{R})$是给定的. 这样一个约束最优问题的传统研究引出了 Lagraùge 函数 $L=x^TAx-\lambda x^Tx$. 于是，它有一个极值的必要条件是

$$0=\nabla L=2(Ax-\lambda x)=0.$$

因此，如果满足 $x^Tx=1$ 的向量 $x\in\mathbf{R}^n$(因而 $x\neq 0$)是 x^TAx 的一个极值点，它必定满足方程 $Ax=\lambda x$，因而 Ax 是 x 的倍数. 这一对 λ，x 称为一个特征值、特征向量偶.

习题

1. 解释(1.0.2)中的约束极值问题为什么一定有一个解，并证明每个实对称矩阵至少有一个实特征值. **提示**：应用 Weierstrass 定理(附录 E)于连续函数 $f(x)=x^TAx$.

2. 设 $A\in M_n(\mathbf{R})$是实对称矩阵($A^T=A$). 证明，在 $x^Tx=1$ 的条件下，x^TAx 的极大值问题的解是 A 的最大特征值.

1.1 特征值-特征向量方程

1.1.1 记号 我们用 $M_n(\mathbf{F})$ 表示域 $\mathbf{F}$ 上的 $n\times n$ 矩阵，通常 $\mathbf{F}$ 取实数域 $\mathbf{R}$ 或复数域 $\mathbf{C}$. 所讨论的问题几乎常常是一些适合于复数矩阵的情形，这时 $M_n(\mathbf{C})$ 简记作 M_n. 对于复数矩阵的一般性质不感兴趣的读者，无论用实数代替复数来阐述什么内容，都很少在表述中，在代数中或在实际中做出本质区别. 但是应注意，常常在讨论多项式的根和其他与"较大"的复数域有关的灵活性问题时，$\mathbf{R}$ 与 $\mathbf{C}$ 之间存在着较大的差别. 通常，最好把实数矩阵看成具有特定元的复数矩阵. 我们知道，有 n 个实分量(相应地，复分量)的所有向量组成的集合(向量空间)用 $\mathbf{R}^n$(相应地 $\mathbf{C}^n$)表示，并且都把它们看作列向量. 最后，$A=[a_{ij}]\in M_n(\mathbf{F})$ 的转置(0.2.5)是矩阵 $[a_{ji}]\in M_n(\mathbf{F})$，记作 A^T，而当 $\mathbf{F}\subseteq\mathbf{C}$ 时，Hermite 伴随是 A 的共轭转置 $[\bar{a}_{ji}]$，记作 A^*. 类似地，如果 $x\in\mathbf{F}^n$，则 x^T 表示与 x 有相同分量的行向量，而当 $\mathbf{F}\subseteq\mathbf{C}$ 时，x^* 表示其分量为 x 的相应分量取复共轭后的行向量. 这里，"$\cdot$"上加一杠 $\bar{\cdot}$ 表示一个复纯量的复共轭(见附录 A)，或者表示一个向量或矩阵按分量取复共轭.

34

矩阵 $A\in M_n$ 可看作从 $\mathbf{C}^n$ 到 $\mathbf{C}^n$ 的线性变换(对于 $\mathbf{C}^n$ 的某个给定的基)，不过把它看作数的一个阵列也是有用的. A 的这两个概念是相互影响的，数的阵列能告诉有关线性变换的信息，而这正是矩阵理论的实质和应用的关键. 或许，矩阵理论中最重要的概念应该是与 A 相关联的 n 个数的集合 $\sigma(A)$，这就是 A 的特征值集合.

1.1.2 定义 设 $A\in M_n$，$x\in\mathbf{C}^n$，考虑方程

$$Ax=\lambda x,\quad x\neq 0,\tag{1.1.3}$$

其中 λ 是纯量. 如果纯量 λ 和非零向量 x 恰好满足这个方程，那么 λ 称为 A 的一个特征值，而 x 称为 A 的属于 λ 的特征向量. 注意，这两个概念不可避免地要成对出现，并且，特征向量不能是零向量.

1.1.4 定义 $A\in M_n$ 的所有特征值 $\lambda\in\mathbf{C}$ 的集合称为 A 的谱，记作 $\sigma(A)$. A 的谱半径是非负实数 $\rho(A)=\max\{|\lambda|:\lambda\in\sigma(A)\}$. 这正是包含 A 的所有特征值的、圆心在复平面原点的最小圆盘的半径.

练习 如果 x 是 A 的、属于 λ 的特征向量，证明 x 的任一非零纯量倍数也是特征向量.

且不说特征值和特征向量有无其他用场，仅从代数上看，它们也有意义，因为，根据(1.1.3)，特征向量是这样一些向量，将它们乘以 A 后有非常简单的形式——同乘以一个纯量(特征值)效果一样.

35

例 考虑矩阵

$$A=\begin{bmatrix}7&-2\\4&1\end{bmatrix}\in M_2.$$

因为

$$A\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}3\\6\end{bmatrix}=3\begin{bmatrix}1\\2\end{bmatrix},$$

于是，有 $3\in\sigma(A)$ 及相应的特征向量 $\begin{bmatrix}1\\2\end{bmatrix}$. 同时，$5\in\sigma(A)$. 可以求得相应于特征值 5 的特征向量.

我们知道，多项式

$$p(t)=a_kt^k+a_{k-1}t^{k-1}+\cdots+a_1t+a_0$$

在矩阵 $A\in M_n$ 处取值是有明确定义的，这是因为可以自乘方阵得到一个正整数幂，并且可以作出同阶矩阵的线性组合，于是，

$$p(A)=a_kA^k+a_{k-1}A^{k-1}+\cdots+a_1A+a_0I. \tag{1.1.5}$$

请注意一个有用的事实，通过多项式的关系得到的与 $A\in M_n$ 相关的矩阵与 A 有相同的特征向量；它的诸特征值与 A 的特征值有简单的关系.

1.1.6 定理　设 $p(\cdot)$ 是给定的多项式. 如果 λ 是 $A\in M_n$ 的特征值，而 x 是相应的特征向量，那么 $p(\lambda)$ 是矩阵 $p(A)$ 的特征值，并且 x 是属于 $p(\lambda)$ 的特征向量.

证明： 考虑 $p(A)x$. 首先，

$$p(A)x\equiv a_kA^kx+a_{k-1}A^{k-1}x+\cdots+a_1Ax+a_0x.$$

其次，反复应用特征值-特征向量方程便有 $A^jx=A^{j-1}Ax=A^{j-1}\lambda x=\lambda A^{j-1}x=\cdots=\lambda^jx$. 因此，

$$p(A)x=a_k\lambda^kx+\cdots+a_0x=(a_k\lambda^k+\cdots+a_0)x=p(\lambda)x. \qquad \square$$

练习　如果 $\sigma(A)=\{-1,2\}$，$A\in M_2$，$\sigma(A^2)$ 是什么？

练习　如果 $D=\mathrm{diag}(d_1,d_2,\cdots,d_n)$ 是对角矩阵(0.9.1)，$\sigma(D)$ 是什么？给出每一个特征值的相应的特征向量. **提示：** 考虑标准基向量 e_i，$i=1,2,\cdots,n$.　[36]

1.1.7 论断　矩阵 $A\in M_n$ 是奇异的，当且仅当 $0\in\sigma(A)$.

证明： 矩阵 A 是奇异的，当且仅当对某个 $x\neq0$，$Ax=0$. 这个关系式成立，当且仅当对某个 $x\neq0$，$Ax=0\cdot x$，即当且仅当 $\lambda=0$ 是特征值. $\square$

习题

1. 假定 $A\in M_n$ 是非奇异的，根据(1.1.7)，这等价于说 A 没有等于零的特征值. 如果 $\lambda\in\sigma(A)$，证明 $\lambda^{-1}\in\sigma(A^{-1})$. 如果 $Ax=\lambda x$，且 $x\neq0$，给出 A^{-1} 的属于 λ^{-1} 的一个特征向量.

2. 如果 $A\in M_n$ 的每一行的各分量之和(简称为行和)是 1，证明 $1\in\sigma(A)$. **提示：** 考虑向量 $e=[1,1,\cdots,1]^T$，然后说明，A 的行和都相等，当且仅当 e 是 A 的特征向量. 如果 A 是非奇异的，证明 A^{-1} 的行和也是 1. 给定多项式 $p(t)$，证明 $p(A)$ 的行和都相等. 它等于什么？

3. 设 $A\in M_n(\mathbf{R})$，如果 λ 是 A 的一个实特征值，且 $Ax=\lambda x$，$0\neq x\in\mathbf{C}^n$，设 $x=\xi+i\eta$，其中 ξ，$\eta\in\mathbf{R}^n$ 是 x 的实部和虚部。证明 $A\xi=\lambda\xi$ 和 $A\eta=\lambda\eta$；由此推出存在 A 的属于 λ 的实特征向量。ξ 和 η 都是 A 的特征向量吗？可能有 A 的属于一个复的非实特征值的实特征向量吗？

4. 考虑分块对角矩阵(0.9.2)

$$A=\begin{bmatrix}A_{11} & 0\\ 0 & A_{22}\end{bmatrix},\quad A_{ii}\in M_{n_i}.$$

证明 A 的特征值由 A_{11} 的特征值和 A_{22} 的特征值组成. **提示**: 先用 A_{11} 和 A_{22} 的特征向量表示 A 的特征向量.

5. 设 $A\in M_n$, 如果 $A^2=A$, 就称 A 为*幂等矩阵*, 证明幂等矩阵的每一个特征值是 0 或 1.

6. 设 $A\in M_n$, 如果对于某个正整数 q, $A^q=0$, 就称 A 为*幂零矩阵*, 上述最小的 q 称为*幂零指标*. 证明幂零矩阵的所有特征值都是 0. 顺便给出一个其特征值都是 0 的非零矩阵的例子.

37

7. 我们将看到, 在集中要讨论的有限维结构中, 每个复的或实的方阵都有一个复特征值. 但是, 对于一个无限维向量空间上的线性变换, 就可能没有任何特征值. 设 V 是由诸复数的形式无限序列组成的向量空间:

$$V=\{(a_1,a_2,\cdots,a_k,\cdots):a_i\in\mathbf{C},i=1,2,\cdots\},$$

并且定义 V 上的线性变换 S 为

$$S(a_1,a_2,\cdots)=(0,a_1,a_2,\cdots).$$

这个线性变换有时称为*移位算子*, 验证 S 是线性变换, 并证明 S 没有任何特征值. **提示**: 证明, 如果一个向量是特征向量, 它的所有分量必须相同, 而这个公共值只能是 0. 因此, 所给出的向量必须是零向量, 它不可能是特征向量.

8. 设矩阵 $A\in M_n$, 如果 $A^*=A$(见 0.2.5), 就称 A 为 Hermite *矩阵*, 如果 A 是 Hermite 的, 证明 A 的所有特征值都是实数. **提示**: 设 $\lambda\in\sigma(A)$ 是任意的, 并且设 x 是相应的特征向量, 于是(1.1.3)推出 $x^*Ax=\lambda x^*x$. 但是 $\overline{x^*Ax}=x^*A^*x=x^*Ax$, 所以 x^*Ax 是实数. 因为 x^*x 是正的, 所以 $\lambda=x^*Ax/x^*x$ 也是实数.

1.2 特征多项式

关于 $A\in M_n$ 的特征值, 一个自然要问到的问题是: A 有多少特征值? 可以怎样来描述它们的特征?

特征值-特征向量方程(1.1.3)可以等价地改写成

$$(\lambda I-A)x=0,\quad x\neq 0. \tag{1.2.1}$$

因而, $\lambda\in\sigma(A)$, 当且仅当 $\lambda I-A$ 是奇异方阵, 即

$$\det(\lambda I-A)=0. \tag{1.2.2}$$

1.2.3 定义 $A\in M_n$ 的*特征多项式*定义为

$$p_A(t)=\det(tI-A),$$

把它看作 t 的形式多项式.

注意 用 t 作为特征多项式的形式变元, 为的是把它和一般的特征值或多项式的零点 λ 区别开来. 在其他地方, 有时用同一个符号表示它们.

1.2.4 论断 如果 $A\in M_n$, 则特征多项式 $p_A(\cdot)$ 的次数为 n, 并且 $p_A(t)=0$ 的根的集合就是 $\sigma(A)$.

38

证明：$p_A(\cdot)$有次数 n 可以归纳地从 $\det(tI-A)$的 Laplace 展开式推出：当行列式展开时，$tI-A$ 的每一行仅提供 t 的一次幂．第二个论断与(1.1.3)和(1.2.2)等价．□

练习　证明，$\det(A-tI)=0$ 与 $\det(tI-A)=0$ 有相同的根，并且证明，$\det(A-tI)=(-1)^n\det(tI-A)$．因此，特征多项式又可以(并且有时就)定义为 $\det(A-tI)$．证明所定义的特征多项式保证 t^n 的(首项)系数总是$+1$．

练习　如果 $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$，证明 $p_A(t)=t^2-(a+d)t+(ad-bc)$，且

$$\sigma(A)=\left\{\frac{a+d\pm\sqrt{(a-d)^2+4bc}}{2}\right\}.$$

设 $A\in M_2(\mathbf{R})$，证明，如果 $bc\geqslant 0$，则 A 的特征值是实数．此外，它们是实数，当且仅当 $(a-d)^2+4bc\geqslant 0$，如果不是实数，则成复共轭对出现．最后证明，如果$(a-d)^2+4bc\neq 0$，则特征值不相同．

在某些一般的情形，矩阵的特征值是容易看出来的．最常见的是因矩阵的形状而使行列式容易计算的情形．其中包括对角矩阵、三角矩阵和一些其他特殊情形．

练习　证明，如果 $T\in M_n$ 是三角矩阵

$$T=\begin{bmatrix} t_{11} & \cdots & t_{1n} \\ & \ddots & \vdots \\ 0 & & t_{nn} \end{bmatrix},$$

那么 $\sigma(T)=\{t_{11}, t_{22}, \cdots, t_{nn}\}$，即 T 的诸对角元之集．

练习　设矩阵 $J_n\in M_n$ 的每个元都等于 1：

$$J_n=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \ddots & & \vdots \\ 1 & \cdots & & 1 \end{bmatrix}.$$

J_2 的特征值是什么？证明 0(出现两次)和 3 是 J_3 仅有的特征值．由此类推，J_n 的特征值是什么？**提示**：考虑向量 $e=[1, 1, \cdots, 1]^T$．

[39]

练习　确定矩阵

$$A=\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$$

的所有特征值和相应的特征向量．**提示**：利用前一个练习，并写出 $A=4I-J_3$．

1.2.5 定义　从(0.7.1)可知，$A\in M_n$ 的$k\times k$ 主子矩阵是位于有相同指标的k 个行和k 个列的子矩阵，而 $k\times k$ 主子式是这个主子矩阵的行列式．$A=[a_{ij}]$有$\binom{n}{k}$个不同的 $k\times k$ 主子式，用 $E_k(A)$表示这些子式的和．特别地，$E_1(A)=\sum_{i=1}^{n}a_{ii}$ 称为A 的迹，通常记作 tr A 或 trace A．注

意，$E_n(A)=\det A$.

练习 如果 $A\in M_2$，证明 $p_A(t)=t^2-(\mathrm{tr}\,A)t+\det A$，且 $\sum_{\lambda\in\sigma(A)}\lambda=\mathrm{tr}\,A$ 和 $\prod_{\lambda\in\sigma(A)}\lambda=\det A$.

一个基本的，但并非显易的事实是，复系数 n 次多项式在复数范围内恰好有 n 个零点(重零点按重数计算)，称这个事实为代数基本定理(附录 C)．由此，可得到如下重要论断.

1.2.6 论断 每个矩阵 $A\in M_n$ 在复数范围内恰好有 n 个特征值(重特征值按重数计算).

注意 这里，当提及 $A\in M_n$ 的一个特征值的"重数"时，就可简单地理解为 λ 作为特征多项式 $p_A(\cdot)$一个零点所出现的次数．更全面地讨论特征值的重数将放到(1.4)节，不过，知道多项式的各阶导数与该多项式的一个零点的重数之间有一定的关系是有用的．多项式 $p(t)$有 λ 作为 $k\geqslant1$ 重零点，当且仅当 $p(t)$可写成 $p(t)=(t-\lambda)^kq(t)$的形式，其中 $q(t)$是使 $q(\lambda)\neq0$ 的多项式．微分这个恒等式就得到 $p'(t)=k(t-\lambda)^{k-1}q(t)+(t-\lambda)^kq'(t)$，并且从这个表示式可以看出，$p'(\lambda)=0$，当且仅当 $k>1$．如果 $k>1$，$p''(t)=k(k-1)(t-\lambda)^{k-2}q(t)+$每一项含有一个因式$(t-\lambda)^m$的多项式，其中 $m\geqslant k-1$．于是，$p''(\lambda)=0$，当且仅当 $k>2$．重复上述计算便可证明，λ 是 $p(t)$的 k 重零点，当且仅当 $p(\lambda)=p'(\lambda)=\cdots=p^{(k-1)}(\lambda)=0$，而 $p^{(k)}(\lambda)\neq0$.

[40]

1.2.7 例 命题(1.2.6)与以下事实密切相关：复数域是代数闭域，也就是，每个系数在该域中的 n 次多项式在该域中有 n 个零点．对于其他域上的矩阵，例如实数域或有理数域，一般几乎不能说出一个矩阵在该域上有多少特征值．但是，再看一看(1.1)节的习题 8，它却是能说出特征值的某些情况的一个例子．在任意域的情形，一个矩阵也可能根本没有不同的特征值．矩阵

$$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\tag{1.2.7a}$$

的所有元尽管都是实数，但它没有实特征值，矩阵

$$\begin{bmatrix}1 & 1 & & & 0\\ & 1 & 1 & & \\ & & 1 & 1 & \\ & & & \ddots & \ddots \\ & & & & 1\\ 0 & & & & 1\end{bmatrix}\tag{1.2.7b}$$

不管它是几阶，也只有一个特殊的特征值(n 重特征值 1).

练习 验证(1.2.7)中的论断.

练习 如果 $A\in M_n(\mathbf{R})$，且 n 为奇数，证明 A 至少有一个实特征值．**提示**：一个实系数多项式的任何非实复零点，必须成共轭对出现，并且注意到，如果 $A\in M_n(\mathbf{R})$，那么 $p_A(\cdot)$有实系数.

根据(1.2.6)，可以把 $A\in M_n$ 的特征值排成

$$\lambda_1,\lambda_2,\cdots,\lambda_n,$$

其中的顺序是任意的，并且按其重数重复这些特征值．于是，因为(1.2.4)，我们得知

$$p_A(t) = (t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n). \tag{1.2.8}$$

1.2.9 定义 n 个数 λ_1，…，λ_n，$k\leqslant n$，的 k 次初等对称函数是

41

$$S_k(\lambda_1,\cdots,\lambda_n) = \sum_{1\leqslant i_1<\cdots<i_k\leqslant n}\prod_{j=1}^{k}\lambda_{i_j},$$

它是所有 $\binom{n}{k}$ 个取自 λ_1，…，λ_n 的不同项的 k 次乘积之和.

例如，$S_1(\lambda_1,\cdots,\lambda_n)=\lambda_1+\cdots+\lambda_n$ 是诸 λ_i 的和，而 $S_n(\lambda_1,\cdots,\lambda_n)=\lambda_1\cdots\lambda_n$ 是诸 λ_i 的乘积. 因为(1.2.8)以及 $p_A(t)$ 是用某个行列式定义的，在矩阵 A 的特征值的初等对称函数 $S_k(\lambda_1,\cdots,\lambda_n)$ 与 A 的各 $k\times k$ 主子式的和 $E_k(A)$(1.2.5)之间存在某种关系. 以下两个恒等式是显然的，费点功夫便可验证：

$$(t-\lambda_1)\cdots(t-\lambda_n) = t^n - S_1(\lambda_1,\cdots,\lambda_n)t^{n-1} + S_2(\lambda_1,\cdots,\lambda_n)t^{n-2} - \cdots \pm S_n(\lambda_1,\cdots,\lambda_n), \tag{1.2.10}$$

以及

$$p_A(t) = t^n - E_1(A)t^{n-1} + E_2(A)t^{n-2} - \cdots \pm E_n(A). \tag{1.2.11}$$

练习 验证(1.2.10)和(1.2.11). 前者可以通过计算乘积 $(t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n)$ 中 t^{n-k} 的系数来直接验证，后者可用 Laplace 展开式归纳地验证.

综合(1.2.10)、(1.2.11)和(1.2.8)，有下面的定理.

1.2.12 定理 如果 λ_1，…，λ_n 是 $A\in M_n$ 的特征值，那么

$$S_k(\lambda_1,\cdots,\lambda_n) = E_k(A).$$

A 的特征值的 k 次初等对称函数是 A 的各 $k\times k$ 主子式之和. 特别地

$$\operatorname{tr} A = \sum_{i=1}^{n}\lambda_i$$

和

$$\det A = \prod_{i=1}^{n}\lambda_i.$$

习题

1. 用(1.2.12)验证(1.1.7).

2. 对于矩阵 $A\in M_{m,n}$ 和 $B\in M_{n,m}$[见(0.2.1)]，通过直接计算证明 $\operatorname{tr} AB=\operatorname{tr} BA$. 再用这个事实证明，对 $A\in M_n$ 和非奇异矩阵 $S\in M_n$，$\operatorname{tr} S^{-1}AS=\operatorname{tr} A$. 矩阵 $S^{-1}AS$ 称为 A 的相似矩阵. 上述结果说明，迹是相似不变量. 相似性是下一节的主题，并且将会看到，所有主子式之和 $E_k(A)$ 都是相似不变量. 注意，因为乘法性质，行列式显然是相似不变量.

42

3. 如果 $D\in M_n$ 是对角矩阵，计算特征多项式 $p_D(t)$，并证明 $p_D(D)=0$.

4. 设 $A\in M_n$，设 $A_i=A(\{i\}')\in M_{n-1}$ 是划去 A 的第 i 行和第 i 列后所得到的 A 的主子矩阵，$i=1$，…，n. 证明

$$\frac{\mathrm{d}}{\mathrm{d}t}p_A(t) = \sum_{i=1}^{n}p_{A_i}(t). \tag{1.2.13}$$

5. 回忆前一节的习题 6，证明幂零矩阵的迹为 0. 幂零矩阵的特征多项式是什么?

6. 如果 $\lambda\in\sigma(A)$ 是 $p_A(t)=0$ 的单重根，$A\in M_n$，证明，$\operatorname{rank}(A-\lambda I)=n-1$，但反过来不一定成立[想一想例(1.2.7b)]. **提示**：利用(1.2.13)和 $t=\lambda$ 时 $(d/dt)p_A(t)\neq 0$ 推出，$A-\lambda I$ 的各个 $n-1$ 阶主子矩阵是非奇异的.

7. 用(1.2.12)确定矩阵

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

的特征多项式. 考虑如何利用这个方法计算一般的 $n\times n$ 三对角矩阵(0.9.10)的特征多项式.

8. 如果 $A\in M_n$，$\sigma(A)=\{\lambda_1, \cdots, \lambda_n\}$，假定 $\sigma(A^k)=\{\lambda_1^k, \cdots, \lambda_n^k\}$. 证明对所有正整数 k

$$\operatorname{tr} A^k = \sum_{i=1}^{n}\lambda_i^k.$$

等式右边的和称为 A 的诸特征值的 k 次矩. 由(2.3.1)知，所作假设成立.

9. 直接计算 $S_2(\lambda_1, \cdots, \lambda_6)$，$S_3(\lambda_1, \cdots, \lambda_6)$，$S_4(\lambda_1, \cdots, \lambda_5)$和 $S_5(\lambda_1, \cdots, \lambda_6)$.

10. 设 V 是域 **F** 上的向量空间. 线性变换 T: $V\to V$ 的特征值是纯量 $\lambda\in\mathbf{F}$，使得有一个非零向量 $v\in V$，适合 $Tv=\lambda v$. 证明，如果 **F** 是复数域，且 V 是有限维的，则每个线性变换 T 有一个特征值. 给出例子说明，如果其中一个假设条件(V 的有限维性质或 $\mathbf{F}=\mathbf{C}$)减弱，那么 T 可能没有任何特征值. **提示**：设 $\mathscr{B}$ 是 V 的基，并考虑$[T]_{\mathscr{B}}$.

43

11. 设 $p(t)=a_nt^n+a_{n-1}t^{n-1}+\cdots+a_1t+a_0$，$a_n=1$，是给定的首系数为 1 的多项式，具有零点 λ_1，λ_2，$\cdots$，λ_n(计相重零点). 诸零点的 k 次矩记作 $\mu_k=\lambda_1^k+\lambda_2^k+\cdots+\lambda_n^k$，$k=1, 2, \cdots$. 证明 Newton 恒等式

$$ka_{n-k}+\mu_1a_{n-k+1}+\mu_2a_{n-k+2}+\cdots+\mu_ka_n=0, k=1,2,\cdots,n. \tag{1.2.14}$$

说明为什么诸零点的前几个矩唯一地确定多项式 $p(t)$的诸系数(因而确定诸零点)，反之亦然. **提示**：证明，对某个 $R>0$，如果 $|t|>R$，那么$(t-\lambda_i)^{-1}=t^{-1}+\lambda_it^{-2}+\lambda_i^2t^{-3}+\cdots$，因而

$$f(t)=\sum_{i=1}^{n}(t-\lambda_i)^{-1}=nt^{-1}+\mu_1t^{-2}+\mu_2t^{-3}+\cdots,\quad |t|>R.$$

证明 $p'(t)=p(t)f(t)$，据此，Newton 恒等式以及关于较高次矩的另一恒等式

$$\mu_ka_0+\mu_{k+1}a_1+\cdots+\mu_{n+k-1}a_{n-1}+\mu_{n+k}a_n=0,\quad k=1,2,\cdots$$

可以通过比较系数推出.

12. 设 A，$B\in M_n$ 是给定的矩阵. 证明，A 和 B 的特征值相同，当且仅当 $\operatorname{tr} A^k=\operatorname{tr} B^k$，$k=1, 2, \cdots, n$. **提示**：利用习题 8 和 Newton 恒等式(1.2.14)证明 A 和 B 的特征多项式相同.

1.3 相似性

正如在(1.0)节所指出的，M_n 中的一个矩阵的相似变换对应于 $\mathbf{C}^n$ 上的一个线性变换在另一个基下的表示. 因此研究相似性可看成是研究一个线性变换所固有的性质或它的所有基表示

所共有的性质.

1.3.1 定义 设矩阵 A，$B\in M_n$，如果存在非奇异矩阵 $S\in M_n$，使得

$$B=S^{-1}AS,$$

则称 B 与 A 相似，而变换 $A\to S^{-1}AS$ 称为由相似矩阵 S 确定的相似变换. 关系"B 与 A 相似"有时简记作 $B\sim A$.

[44]

1.3.2 论断 相似是 M_n 上的一个等价关系；即相似是

(a) 自反的：$A\sim A$；

(b) 对称的：$B\sim A$ 推出 $A\sim B$；

(c) 传递的：$C\sim B$ 和 $B\sim A$ 推出 $C\sim A$.

练习 验证(1.3.2).

像任何等价关系一样，相似关系把集合 M_n 划分成互不相交的等价类. 每一个等价类是 M_n 中相似于某个矩阵(该类的代表)的所有矩阵的集合. 在一个等价类中的所有矩阵都相似，而属于两个不同类的矩阵不相似. 由于传递性，在任何一个相似矩阵的有限序列中，第一个矩阵和最后一个矩阵在同一个相似等价类中. 一个至关重要的结果是，任一个等价类中的矩阵共同具有许多重要性质. 其中一些将在这里论述，而关于相似不变量的一个较完整的描述(例如，Jordan 标准形)将放在后面的第 3 章.

1.3.3 定理 设 A，$B\in M_n$. 如果 B 和 A 相似，那么 B 的特征多项式与 A 的相同.

证明：对任意 t，我们有

$$\begin{aligned}p_B(t)&=\det(tI-B)\\&=\det(tS^{-1}S-S^{-1}AS)=\det S^{-1}(tI-A)S\\&=\det S^{-1}\det(tI-A)\det S\\&=(\det S)^{-1}(\det S)\det(tI-A)\\&=\det(tI-A)=p_A(t).\end{aligned}$$

□

1.3.4 推论 如果 A，$B\in M_n$，且 A 与 B 相似，那么它们有相同的特征值(重特征值按重数计算).

1.3.5 例 有相同的特征值是相似的必要条件，但不是充分条件，考虑矩阵

$$\begin{bmatrix}0&1\\0&0\end{bmatrix} \quad 和 \quad \begin{bmatrix}0&0\\0&0\end{bmatrix},$$

每一个都有二重特征值 0，但它们不相似.

练习 证明与零矩阵相似的矩阵只有它本身，然后利用这一事实验证例(1.3.5)中的论断.

[45]

练习 如果矩阵 A，$B\in M_n$ 相似，并且 $q(\cdot)$ 是多项式，证明 $q(A)$ 与 $q(B)$ 相似. 特别地，证明，如果 α 是纯量，那么 $A+\alpha I$，与 $B+\alpha I$ 相似.

练习 如果 A，B，C，$D\in M_n$，且 $A\sim B$ 和 $C\sim D$ 是经同一相似矩阵 S 实现的，证明 $A+C\sim B+D$.

练习 如果 A，$S\in M_n$，且 S 是非奇异矩阵，证明 $E_k(S^{-1}AS)=E_k(A)$，特别是，$\det S^{-1}AS$

$=\det A$ 和 $\operatorname{tr} S^{-1}AS=\operatorname{tr} A$，即行列式，迹和其他的 $k\times k$ 主子式和是相似不变量.

练习 证明秩也是相似不变量：如果 $B\in M_n$ 相似于 $A\in M_n$，那么 $\operatorname{rank} B=\operatorname{rank} A$. **提示**：见(0.4.6).

因为对角矩阵特别简单，又有很好的性质，因此有必要知道，对于哪些矩阵 $A\in M_n$，在 A 的相似等价类中存在一个对角矩阵，即哪些矩阵相似于对角矩阵.

1.3.6 定义 如果矩阵 $A\in M_n$ 与一个对角矩阵相似，那么就说 A *可对角化*. 有时也采用术语*可对角的*.

1.3.7 定理 设 $A\in M_n$. 那么，A 可对角化，当且仅当 A 有 n 个线性无关的特征向量.

证明：如果 A 有 n 个线性无关的特征向量 $x^{(1)}$，…，$x^{(n)}$. 以它们为列作非奇异矩阵 S，通过计算，

$$\begin{aligned}S^{-1}AS &= S^{-1}[Ax^{(1)}Ax^{(2)}\cdots Ax^{(n)}]\\ &= S^{-1}[\lambda_1 x^{(1)}\cdots\lambda_n x^{(n)}] = S^{-1}[x^{(1)}\cdots x^{(n)}]\Lambda\\ &= S^{-1}S\Lambda = \Lambda,\end{aligned}$$

其中

$$\Lambda=\begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix}.$$

而 λ_1，…，λ_n 是 A 的特征值.

反过来，假定存在相似矩阵 S 使得 $S^{-1}AS=\Lambda$ 是对角矩阵. 于是 $AS=S\Lambda$. 这就是说，A 乘 S 的第 i 列(即 AS 的第 i 列)是 Λ 的第 i 个对角元乘 S 的第 i 列(即 $S\Lambda$ 的第 i 列)，或者说，S 的第 i 列是 A 的相应于 Λ 的相应于 Λ 的第 i 个对角元的特征向量. 因为 S 是非奇异的，所以存在 n 个线性无关的特征向量. □

[46]

更注意的是，(1.3.7)的证明原则上是关于对角化一个可对角矩阵的算法：求 A 的各特征值；求相应的各个特征向量(考虑重特征值)，然后把它们排成矩阵 S. 如果诸特征向量线性无关，那么 S 是一个对角化相似矩阵. 但是，我们要着重指出，这只是粗略的分析性解释，不是实际的计算方法.

附注 如果 $A\in M_n$ 可对角化，与 A 相似的任一对角矩阵的各对角元必须是 A 的具有适当重数的特征值. 此外，线性无关的特征向量(它们组成相似矩阵)必须对应具有适当重数的不同的特征值；即，如果 $x^{(1)}$，…，$x^{(n)}$ 是线性无关的特征向量，且 $p_A(t)=(t-\lambda_1)\cdots(t-\lambda_n)$，那么，对诸指标的某个排列 τ，有 $Ax^{(i)}=\lambda_{\tau(i)}x^{(i)}$.

练习 证明矩阵 $A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$ 不能对角化.

其理由是：一方面是因为，如果它可对角化，它将相似于0矩阵，而这是不可能的；另一方面，经计算，除了差一个比例因子以外，只存在一个属于0的特征向量.

练习 如果 A 可对角化，而 $q(\cdot)$ 是一个多项式，证明 $q(A)$ 可对角化. **提示**：$q(SAS^{-1})=$

$Sq(A)S^{-1}$.

练习 如果 $A\in M_n$，且 $\lambda\in\sigma(A)$ 作为 A 的特征值有重数 m，证明，如果 $\operatorname{rank}(A-\lambda I)>n-m$，那么 A 不能对角化.

保证可对角化性质可行的一个简单情形是矩阵的各特征值互不相同. 这个事实的一个重要前提是下述引理，它还有其他用途：

1.3.8 引理 假定 $\lambda_1,\cdots,\lambda_k$ 是 $A\in M_n$ 的两两不相同的特征值，而 $x^{(i)}$ 是相应于 λ_i 的特征向量，$i=1,\cdots,k$. 那么 $\{x^{(1)},\cdots,x^{(k)}\}$ 是线性无关组.

证明：证明实质上是用反证法. 相反，假设 $x^{(1)},\cdots,x^{(k)}$ 是一个线性相关组，那么存在一个等于 0 向量的非平凡线性组合，并且实际上有这样一个线性组合，它的非零系数最少. 假定这个极小的线性相关关系式是

47

$$\alpha_1 x^{(1)}+\alpha_2 x^{(2)}+\cdots+\alpha_r x^{(r)}=0,\quad r\leqslant k.$$

因为所有 $x^{(i)}\neq 0$，有 $r>1$. 为方便起见，可以假定它包含前 r 个向量(如果必要，可重排编号). 同时，还有另一个相关关系式

$$\begin{aligned}A(\alpha_1 x^{(1)}+\cdots+\alpha_r x^{(r)})&=\alpha_1 Ax^{(1)}+\cdots+\alpha_r Ax^{(r)}\\&=\alpha_1\lambda_1 x^{(1)}+\cdots+\alpha_r\lambda_r x^{(r)}=0.\end{aligned}$$

现在用 λ_r 乘第一个关系式，然后从第二个关系式中减去它便得到第三个相关关系式

$$\alpha_1(\lambda_1-\lambda_r)x^{(1)}+\cdots+\alpha_{r-1}(\lambda_{r-1}-\lambda_r)x^{(r-1)}=0,$$

它的非零系数比第一个关系式要少. 因为 $\lambda_i\neq\lambda_r$，$i=1,2,\cdots,r-1$，这最后一个关系式是非平凡的. 这就与第一个相关关系式的极小性假设相矛盾，因而得证. □

1.3.9 定理 如果 $A\in M_n$ 有 n 个互不相同的特征值，那么 A 可对角化.

证明：如果 $\sigma(A)=\{\lambda_1,\cdots,\lambda_n\}$，设 $x^{(i)}$ 是相应于 λ_i 的特征向量. 因为特征值都各不相同，根据(1.3.8)，$\{x^{(1)},\cdots,x^{(n)}\}$ 是线性无关组，因此，再由(1.3.7)可知，A 可对角化.

练习 给出一个可对角化矩阵 $A\in M_n$ 的例子，但它没有互不相同的特征值.

练习 由(0.9.5)想到，置换矩阵 P 是其每一行和每一列中恰有一个分量为 1 的以 0，1 为分量的矩阵. 因而 $P^T=P^{-1}$. 证明 $A\in M_n$ 的一个置换相似重排 A 的诸对角元，然后证明，对任一对角矩阵，存在一个置换相似矩阵，使其对角元可按任意顺序重排，特别是任一重复出现的对角元可相邻地排放在一起.

矩阵 A，$B\in M_n$ 关于乘法一般不交换，但是，如果 A，B 都是对角矩阵，它们总是可交换的. 这后一个结论可以做些推广；在这方面，下面的引理是有益的.

48

1.3.10 引理 设 $A\in M_n$ 和 $B\in M_m$ 是给定的矩阵，且设

$$C=\begin{bmatrix}A&0\\0&B\end{bmatrix}$$

是 A 与 B 的直和，那么，C 可对角化，当且仅当 A 和 B 都可对角化.

证明：如果存在非奇异矩阵 $S_1\in M_n$ 和非奇异矩阵 $S_2\in M_m$，使得 $S_1^{-1}AS_1$ 和 $S_2^{-1}BS_2$ 都是对角矩阵，那么容易验证 $S^{-1}CS$ 是对角矩阵，只要 S 取直和

$$S=\begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}.$$

反之，设 C 可对角化，存在非奇异矩阵 $S\in M_{n+m}$，使 $S^{-1}CS=\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{n+m})$ 是对角矩阵．如果用

$$s_i=\begin{bmatrix} \xi_i \\ \eta_i \end{bmatrix}\in \mathbf{C}^{n+m}, \xi_i\in \mathbf{C}^n, \eta_i\in \mathbf{C}^m,\quad i=1,2,\cdots,n+m$$

表示 $S=[s_1 s_2\cdots s_{n+m}]$，那么，对 $i=1, 2, \cdots, n+m$，$Cs_i=\lambda_i s_i$ 推出 $A\xi_i=\lambda_i\xi_i$ 和 $B\eta_i=\lambda_i\eta_i$ 如果在集合 $\{\xi_1, \cdots, \xi_{n+m}\}$ 中，无关向量少于 n 个，则矩阵

$$[\xi_1\xi_2\cdots\xi_{n+m}]\in M_{n,n+m}$$

的列秩(因而行秩)将小于 n．同理，如果在集合 $\{\eta_1, \cdots, \eta_{n+m}\}$ 中，无关向量少于 m 个，则矩阵

$$[\eta_1\eta_2\cdots\eta_{n+m}]\in M_{m,n+m}$$

的列秩(因而行秩)将小于 m．在其中一种(或两种)情形下，矩阵

$$S=[s_1\cdots s_{m+n}]=\begin{bmatrix} \xi_1 & \cdots & \xi_{n+m} \\ \eta_1 & \cdots & \eta_{n+m} \end{bmatrix}\in M_{n+m}$$

的行秩(因而秩)小于 $n+m$；因为 S 是可逆的，所以这是不可能的．因此，在集合 $\{\xi_1, \xi_2, \cdots, \xi_{n+m}\}$ 中恰有 n 个线性无关的向量，又因为这每一个向量都是 A 的特征向量，所以 A 一定可对角化．同理可证矩阵 B 可对角化． □

1.3.11 定义　我们说两个可对角化矩阵 A，$B\in M_n$ 同时可对角化，指的是存在同一个相似矩阵 $S\in M_n$，使得 $S^{-1}AS$ 和 $S^{-1}BS$ 都是对角矩阵，即，如果存在同一个基，在这个基下，两个线性变换的表示都是对角矩阵．

练习　证明，如果 A，$B\in M_n$ 同时可对角化．那么它们可交换．提示：写出 $A=SDS^{-1}$ 和 $B=SES^{-1}$，D 和 E 都是对角矩阵．然后利用对角矩阵是交换的事实计算 AB 和 BA．这种处理方式会经常用到．

练习　证明，如果 $A\in M_n$ 可对角化，而 λI 是 M_n 中的一个纯量矩阵，那么 A 和 λI 同时可对角化．

1.3.12 定理　设 A，$B\in M_n$ 可对角化．那么，A 和 B 可交换，当且仅当它们同时可对角化．

证明：假定 A 和 B 可交换，在 A 和 B 上同施以一个相似变换使 A 对角化，因而，不失一般性，可以假定 A 是对角矩阵，仍不失一般性，再假定 A 的任一多重特征值相邻地出现在主对角线上．因为 $AB=BA$(上述公共的相似变换不会改变这一关系)，所以有

$$\lambda_i b_{ij}=b_{ij}\lambda_j,$$

其中，$B=[b_{ij}]$，而 $\lambda_1, \cdots, \lambda_n$ 是 A 的各特征值．因为 $(\lambda_i-\lambda_j)b_{ij}=0$，由此可知，只要 $\lambda_i\neq\lambda_j$，就有 $b_{ij}=0$．因此，按上面已经给定的 λ_i 项的顺序，B 是分块对角矩阵：

$$B=\begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_k \end{bmatrix},\tag{1.3.13}$$

其中，对于 A 的每个不同的特征值，有一个子块 B_i. 每个 B_i 是一个方阵，其阶数是与它相应的 A 的特征值的重数. 因为 B 可对角化，根据(1.3.10)，每个 B_i 可对角化. 设 T_i 是使 $T_i^{-1}B_iT_i$ 为对角矩阵的非奇异矩阵. 因为 A 有分块形式

$$A=\begin{bmatrix}\lambda_1 I & & & 0\\ & \lambda_2 I & & \\ & & \ddots & \\ 0 & & & \lambda_k I\end{bmatrix},\tag{1.3.14}$$

[50]

其中每个纯量矩阵 $\lambda_i I$ 与 B_i 同阶，我们看到 $T^{-1}AT$ 与 $T^{-1}BT$ 都是对角矩阵，其中 T 是直和

$$T=\begin{bmatrix}T_1 & & & 0\\ & T_2 & & \\ & & \ddots & \\ 0 & & & T_k\end{bmatrix},\tag{1.3.15}$$

注意，$T_i^{-1}\lambda_i IT_i=\lambda_i I$.

逆命题已包括在前面一个练习中. □

作为本节的结束，把(1.3.12)推广到较大的矩阵集合，并且对不可对角化矩阵的情形给出一个较弱的结果.

1.3.16 定义 矩阵的一个族 $\mathscr{F}\subseteq M_n$ 是矩阵的任一(有限的或无限的)集合，而交换族是其每一对矩阵在乘法下都是可交换的族. 我们称子空间 $W\subseteq \mathbf{C}^n$ 对 $A\in M_n$ 是 A-不变的，是指 $Aw\in W$ 对每个 $w\in W$ 成立；称 W 对族 $\mathscr{F}\subseteq M_n$ 是 $\mathscr{F}$-不变的，是指 W 对每个 $A\in\mathscr{F}$ 是 A-不变的.

注意. 如果 $A\in M_n$，$\mathbf{C}^n$ 的一维 A-不变子空间中的每个非零元素是 A 的特征向量.

练习 设 $A\in M_n$. 如果 W 是维数至少为 1 的 $\mathbf{C}^n$ 的 A-不变子空间，证明在 W 中有 A 的一个特征向量. **提示**：选取 W 的一个基，然后考虑作为 W 上的线性变换 $T: w\to Aw$ 的基表示矩阵. 证明这个矩阵有一个特征值. 要点是：为什么 T 是 W 上的线性变换？

一个重要的结论是下面的引理.

1.3.17 引理 如果 $\mathscr{F}\subseteq M_n$ 是交换族，那么，存在向量 $x\in\mathbf{C}^n$，它是每个 $A\in\mathscr{F}$ 的特征向量.

证明：设 $W\subseteq\mathbf{C}^n$ 是有最小正维数的 $\mathscr{F}$-不变子空间；这样的 W 存在，但未必唯一. 因为 $\mathbf{C}^n$ 本身就是 $\mathscr{F}$-不变的，所以知道有一个 n 维 $\mathscr{F}$-不变子空间. 如果存在 $n-1$ 维 $\mathscr{F}$-不变子空间，那么就要问是否存在 $n-2$ 维 $\mathscr{F}$-不变子空间，等等. 实际上，只要证明 W 中的每个非零向量是每个 $A\in\mathscr{F}$ 的一个特征向量，就完成了引理的证明. 假如上述情形不成立，那么，对某个矩阵 $A\in\mathscr{F}$，并非 W 中每个非零向量都是 A 的特征向量. 但是，因为 W 是 $\mathscr{F}$-不变的，所以它是 A-不变的，因而在 W 中有 $x\neq 0$，使得 $Ax=\lambda x$ 对某个特征值 λ 成立. 定义 $W_0=\{y\in W: Ay=\lambda y\}$，于是 $x\in W_0$，且 $W_0\subseteq W$ 是一个子空间. 因为关于 A 的假设，$W_0\neq W$，因而 W_0 的(正)维数严格小于 W 的维数. 设 $B\in\mathscr{F}$，如果 $x\in W_0$，则有 $Bx\in W$，这是因为 $W_0\subseteq W$ 且 W 是 $\mathscr{F}$-不变的. 但是另一方面，因为 $\mathscr{F}$ 是交换族，$A(Bx)=(AB)x=(BA)x=B(Ax)=B(\lambda x)=$

[51]

$\lambda(Bx)$，因而得出 $Bx\in W_0$. 由此可知，W_0 是 $\mathscr{F}$-不变的. 但因为 W_0 有严格低于 W 的正维数，这就产生了矛盾. 证毕. □

引理(1.3.17)是关于任意基数的交换族的. 特别是，如果 $\mathscr{F}=\{A, B\}$ 是只有两个矩阵的族，那就是说，任一对交换矩阵有一个公共的特征向量. 定理(1.3.12)是说，如果 A 和 B 不仅可交换而且每一个也都可对角化，那么它们同时可对角化. 我们的下一个结果要证明，关于两个可对角化矩阵的交换族的上述性质不是它所特有的；这个结论可以推广到具有任意基数的族.

1.3.18 定义 同时可对角化的族 $\mathscr{F}\subset M_n$ 是这样一个族，关于这个族，存在同一个非奇异矩阵 $S\in M_n$，使得对每一个 $A\in\mathscr{F}$，$S^{-1}AS$ 是对角矩阵.

1.3.19 定理 设 $\mathscr{F}\subset M_n$ 是由可对角化矩阵组成的族. 那么，$\mathscr{F}$ 是交换族，当且仅当它是同时可对角化的族.

证明： 如果 $\mathscr{F}$ 同时可对角化，那么，根据前面的练习，它是交换族. 对 n 作归纳法来证明其逆命题. 如果 $n=1$，就没有什么可证的了，因为每个族既是交换的，也是对角的. 假设 $n\geq 2$，并且假定，对 $k=1, 2, \cdots, n-1$，关于满足假设的所有 $k\times k$ 矩阵族，结论已经证明. 如果 $\mathscr{F}$ 的每个矩阵是纯量矩阵，那就无须证明，因此，可以假定，$A\in\mathscr{F}$ 是某个具有特征值 λ_1，λ_2，$\cdots$，λ_k(其中至少有两个不相同，$2\leq k\leq n$)的 $n\times n$ 可对角化矩阵，还假定，对每个矩阵 $B\in\mathscr{F}$，$AB=BA$，且每个 $B\in\mathscr{F}$ 可对角化. 采用与(1.3.12)中相同的论证，可以把情况简化为，A 实际上是对角矩阵，A 的任一多重特征值相邻地出现，且特征值的顺序是固定的，即 A 有形式(1.3.14). 因为每个 $B\in\mathscr{F}$ 与 A 交换，(1.3.12)中证明，每个 $B\in\mathscr{F}$ 有阶数为 $n-1$ 或小于 $n-1$ 的矩阵直和形式(1.3.13). 在(1.3.13)中子块的阶数和位置完全由 A 的诸特征值的重数和顺序所确定，因此，对于所有 $B\in\mathscr{F}$，它们都是相同的. 因为，所有矩阵 $B\in\mathscr{F}$ 都可交换(不只是与 A)，且每个 $B\in\mathscr{F}$ 有一个直和形式(1.3.13)，所以，$\mathscr{F}$ 中任一矩阵的 k 个直和被加子块中的每一个都是 $\mathscr{F}$ 的其他每个矩阵的相应子块可交换，并且，根据(1.3.10)，这每个子块都可对角化. 由归纳假设，存在 k 个相应阶数的相似矩阵 T_1，T_2，$\cdots$，T_k，它们中的每一个都使 $\mathscr{F}$ 中的每个矩阵的对应子块对角化. 正如(1.13.15)中的直和那样，直和 $T_1\oplus T_2\oplus\cdots\oplus T_k$ 使 $\mathscr{F}$ 中的每个矩阵对角化. □

附注 与这一节相关的两个重要问题将推迟到第3章讨论：(1) 给定 A，$B\in M_n$，如何确定 A 是否与 B 相似？这是促成求相似下的标准形的动机. (2) 不计算已知矩阵 $A\in M_n$ 的特征向量，我们如何判别它是否可对角化？

作为交换性的最后一个附注，我们注意到，虽然 AB 与 BA 未必是相同的矩阵(并且即使两者都有定义，它仍未必是同阶的)，但是从它们的特征值来看，几乎是相同的，如果 A 和 B 都是方阵，AB 和 BA 恰有相同的特征值.

1.3.20 定理 假定 $A\in M_{m,n}$，$B\in M_{n,m}$，且 $m\leq n$. 那么 BA 与 AB 有相同的特征值(重特征值按重数计算)，再附加 $n-m$ 个等于0的特征值；即 $p_{BA}(t)=t^{n-m}p_{AB}(t)$. 如果 $m=n$，且 A 或 B 至少有一个非奇异，那么 AB 与 BA 相似.

证明：考虑以下两个涉及 M_{m+n} 中的分块矩阵的恒等式：

$$\begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix}\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}=\begin{bmatrix} AB & ABA \\ B & BA \end{bmatrix},$$

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}\begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix}=\begin{bmatrix} AB & ABA \\ B & BA \end{bmatrix}.$$

因为分块矩阵

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}\in M_{m+n}$$

53

非奇异（它的所有特征值是 +1），得出

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}^{-1}\begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix}\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}=\begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix},$$

即两个 $(m+n)\times(m+n)$ 矩阵

$$C_1=\begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix} \quad 和 \quad C_2=\begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix}$$

相似．C_1 的特征值是 AB 的特征值再加上 n 个零．C_2 的特征值是 BA 的特征值再加上 m 个零．因为根据(1.3.4)，C_1 与 C_2 的特征值相同（计相重特征值），所以定理的主要论断已经证明．最后一个论断可从以下结果推出：如果 A 是非奇异的，且 $m=n$，那么 $AB=A(BA)A^{-1}$．

习题

1．如果 A，$B\in M_n$，且 A 与 B 可交换，证明 A 和关于 B 的任一多项式可交换．

2．设 A，$B\in M_n$，且有 $\sigma(A)=\{\lambda_1,\cdots,\lambda_n\}$ 和 $\sigma(B)=\{\mu_1,\cdots,\mu_n\}$．如果 A 和 B 可对角化，且可交换，证明，存在 $1,\cdots,n$ 的某个排列 $i_1,\cdots,i_n$，使得 $A+B$ 的特征值是

$$\lambda_1+\mu_{i_1},\lambda_2+\mu_{i_2},\cdots,\lambda_n+\mu_{i_n}.$$

3．如果 $A\in M_n$ 和 $A=S^{-1}DS$，$D=\text{diag}(d_1,\cdots,d_n)$，且 $p(\cdot)$是多项式，证明 $p(A)=S^{-1}p(D)S$ 和 $p(D)=\text{diag}(p(d_1),\cdots,p(d_n))$．只要能使 A 对角化，这就提供了计算 $p(A)$的一个简便方法．

4．给出两个交换矩阵不可同时对角化的例子．这与定理(1.13.12)矛盾吗？

5．如果 $A\in M_n$ 有互不相同的特征值，且与给定的矩阵 $B\in M_n$ 可交换，证明 B 是次数至多为 $n-1$ 的关于 A 的多项式．**提示**：采用在定理(1.3.12)的证明中使用过的方法，证明 B 和 A 一定同时可对角化．然后想到，给定互不相同的数 $\alpha_1,\cdots,\alpha_n$ 和 $\beta_1,\cdots,\beta_n$，存在一个次数至多为 $n-1$ 的(Lagrange 插值)多项式 $p(\cdot)$，使得 $p(\alpha_i)=\beta_i$．见(0.9.11)．

6．如果 $A\in M_n$ 可对角化，考虑特征多项式 $p_A(t)$，证明 $p_A(A)$是零矩阵．

7．设矩阵 A，$B\in M_n$，如果 $A^2=B$，就称 A 是 B 的平方根．证明 M_n 中的每个可对角化矩阵有一个平方根．

54

8．如果 A，$B\in M_n$，且至少有一个有互不相同的特征值（关于另一个，甚至连它可对角化都没有假设），证明，A 和 B 可交换，当且仅当它们同时可对角化．**提示**：充分性的证明是容易的；至于必要性，试图采取如下形式的论证来作为(1.3.12)所采用的方法的一个补充．假定

B 有互不相同的特征值，$\lambda\in\sigma(B)$，且 $Bx=\lambda x$ 及 $x\neq 0$. 于是 $B(Ax)=A(Bx)=A\lambda x=\lambda Ax$，由此推出 Ax 也是 B 的属于 λ 的特征向量. 因为不可能存在两个这样的线性无关的向量(因为 λ 是单重的)，所以 Ax 必须是 x 的 μ 倍：即 $Ax=\mu x$. 因此，B 的每个特征向量也是 A 的特征向量，并且使 B 对角化的这些特征向量所组成的同一个矩阵也使 A 对角化. 有关这同一个命题的其他处理方法，见习题 12 和 13.

9. 对定理(1.3.20)的下述另一个证明作详细的论述. (a)首先，假定 $A, B\in M_n$，且其中至少有一个是非奇异的. 证明 AB 相似于 BA；因而 AB 和 BA 的特征多项式相同. **提示**：若 A 是非奇异的，则 $BA=A^{-1}(AB)A$. 此时，$\sigma(AB)=\sigma(BA)$. (b)考虑奇异矩阵 $A=\begin{bmatrix}1&0\\0&0\end{bmatrix}$和 $B=\begin{bmatrix}0&0\\1&0\end{bmatrix}$. 证明 AB 与 BA 不相似，但它们有相同的特征值. (c)证明，若 $A, B\in M_n$，则 AB 与 BA 有相同的特征值，包括重特征值. **提示**：考虑下面的分析论证. 对所有充分小的 $\varepsilon>0$，$A_\varepsilon\equiv A+\varepsilon I$ 是非奇异的；因而 $A_\varepsilon B$ 与 BA_ε 相似，故 $A_\varepsilon B$ 与 BA_ε 有相同的特征多项式. 如果我们现在令 $\varepsilon\to 0$，取极限不能保证其相似性，但其特征多项式仍然相等，这是因为 $p_{A_\varepsilon B}(t)=\det(tI-A_\varepsilon B)$ 连续地依赖 ε. 因此 AB 与 BA 有相同的特征多项式，因而有相同的特征值，包括重特征值. (d)最后，若 $A\in M_{m,n}$，$B\in M_{n,m}$，证明，AB 与 BA 有相同的特征值，包括重特征值，但不包括 BA 另有的 $n-m$ 个为 0 的特征值(假定 $n>m$)；等价地，$p_{BA}(t)=t^{n-m}p_{AB}(t)$. **提示**：从 A(添加若干 0 行)以及 B(添加若干 0 列)作两个新的 $n\times n$ 矩阵，利用最后一个结果，把两个新的(经过适当分块的)矩阵乘积与原有的两个乘积进行比较.

10. 利用(1.3.8)证明下述推广：设 $A\in M_n$ 已知，且 $\lambda_1, \cdots, \lambda_k$ 是 A 的互不相同的特征值. 对于每个 $i=1, 2, \cdots, k$，假定 $\{x_1^{(i)}, x_2^{(i)}, \cdots, x_{n_i}^{(i)}\}$ 是 A 的相应于特征值 λ_i 的 $n_i\geq 1$ 个特征向量的无关组. 证明，诸集合之并 $\{x_1^{(1)}, x_2^{(1)}, \cdots, x_{n_1}^{(1)}\}\cup\cdots\cup\{x_1^{(k)}, x_2^{(k)}, \cdots, x_{n_k}^{(k)}\}$ 是一个无关组. **提示**：如果某个线性组合是零，比如

[55]

$$0=\sum_{i=1}^{k}\sum_{j=1}^{n_i}c_{ij}x_j^{(i)}=\sum_{i=1}^{k}y^{(i)},$$

利用(1.3.8)证明每个 $y^{(i)}=0$.

11. 对引理(1.3.17)的下述另一个更具构造性的证明作详细论述. (a) 证明，若 $A, B\in M_n$ 可交换，则它们有一个公共特征向量. **提示**：设 x 是 A 的一个特征向量，$Ax=\lambda x$，$x\neq 0$，然后考虑序列 $x, Bx, B^2x, B^3x, \cdots$. 这个序列中一定有一个元素与它前面的元素线性相关，取最靠前的这种元素，如 B^kx，所以 $S=\mathrm{Span}\{x, Bx, B^2x, \cdots, B^{k-1}x\}$ 是 B 的一个不变子空间，因而存在某个非零 $y\in S$ 使得 $By=\mu y$. 但是 $AB^jx=B^jAx=B^j\lambda x=\lambda B^jx$，因而 S 中的每个非零向量也是 A 的特征向量. (b) 若 $\mathcal{F}=\{A_1, A_2, \cdots, A_m\}$ 是一个有限交换族，用归纳法证明，对所有 A_i 有一个公共的特征向量. **提示**：若 $y\neq 0$ 是 $A_1, A_2, \cdots, A_{m-1}$ 的一个公共特征向量，像(a)中那样，考虑序列 $y, A_my, A_m^2y, A_m^3y, \cdots$. (c) 若 $\mathcal{F}\subset M_n$ 是一个没有有限基数的交换族. 注意到在 $\mathcal{F}$ 中不可能有多于 n^2 个线性无关矩阵. 选一个极大无关组再利用(b)证明，这个有限组的公共特征向量是 $\mathcal{F}$ 的所有元素的公共特征向量.

12. 如果 $A=\operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)\in M_n$ 有 n 个互不相同的对角元，用定理(1.3.12)的证明思想证明，对某个 $B\in M_n$，$AB=BA$，当且仅当 B 本身是对角矩阵(但不必具有不同的对角元).

13. 假设 $A\in M_n$ 有 n 个互不相同的特征值. 如果对某个 $B\in M_n$，$AB=BA$，证明 B 可对角化，且 A 和 B 同时可对角化. **提示：**如果 $A=S\Lambda S^{-1}$，Λ 是对角矩阵，证明 Λ 与 $S^{-1}BS$ 可交换，然后利用习题 12.

14. 把习题 13 的结果推广到交换族 $\mathcal{F}\subset M_n$，这个族至少包含一个具有 n 个互不相同的特征值的矩阵. 试将这个结果与假定族的所有成员都是可对角化的定理(1.3.19)作一比较，这是一个较强的结果吗？

15. 考虑分块对角矩阵 $A=\operatorname{diag}(\lambda_1 I_1, \lambda_2 I_2, \cdots, \lambda_k I_k)\in M_n$，其中，$I_j\in M_{n_j}$，如果 $i\neq j$，则 $\lambda_i\neq\lambda_j$，且 $n_1+n_2+\cdots+n_k=n$. 证明，对某个 $B\in M_n$，$AB=BA$，当且仅当矩阵 B 有分块对角矩阵形式 $B=\operatorname{diag}(B_1, B_2, \cdots, B_k)$，其中，$B_j\in M_{n_j}$，$j=1, 2, \cdots, k$. 这个结果与习题 12 有何关系？

16. 设 A，$B\in M_n$，且假定 A 或 B 非奇异. 如果 AB 可对角化，证明 BA 也可对角化. 考虑 $A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$和 $B=\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}$，说明，如果 A 和 B 都是奇异的，上述结论未必成立. [56]

1.4 特征向量

迄今为止，已经着重讨论了 $A\in M_n$ 的特征值，也讨论了相应的特征向量. 特征向量不仅在对角化中起重要作用，而且在各种应用中也有它们的用场. 为此还要较深入地讨论特征向量. 不过先从关于特征值的另一个论断开始.

1.4.1 论断 设 $A\in M_n$，那么，(a) A^T 与 A 有相同的特征值(重特征值按重数计算). (b) A^* 的特征值是 A 的特征值的复共轭(重特征值按重数计算).

证明：因为 $\det(tI-A^T)=\det(tI-A)^T=\det(tI-A)$，所以 $p_{A^T}(t)=p_A(t)$. 因而(a)得证. 类似地，$\det(\bar{t}I-A^*)=\det[(tI-A)^*]=\overline{\det(tI-A)}$，由此推知 $p_{A^*}(\bar{t})=\overline{p_A(t)}$. 因而(b)得证. □

练习 如果 x，$y\in \mathbf{C}^n$ 都是 A 的相应于特征值 λ 的特征向量，证明 x 和 y 的任一非零线性组合也是相应于 λ 的特征向量. 由此得出，属于某个特征值 $\lambda\in\sigma(A)$ 的所有特征向量连同零向量组成的集合是 $\mathbf{C}^n$ 的一个子空间.

练习 说明在上述练习中所描述的子空间恰好是 $A-\lambda I$ 的零空间.

1.4.2 定义 设 $A\in M_n$. 对于给定的 $\lambda\in\sigma(A)$，满足 $Ax=\lambda x$ 的所有向量 $x\in\mathbf{C}^n$ 的集合称为 A 的相应于特征值 λ 的*特征空间*，注意，这个特征空间的每个非零元素是 A 的相应于 λ 的一个特征向量.

练习 证明，A 的相应于特征值 λ 的特征空间是 A-不变子空间，但是，反之不成立. 证明一个极小的 A-不变子空间(不真包含较低维的非平凡 A-不变子空间)是单独由 A 的一个特征向量生成的. **提示：**运用(1.3.17)之前的练习.

如果知道 $A\in M_n$ 的一个特征值，一个计算相应特征向量的方法是解线性方程组 [57]

$$(A-\lambda I)x=0.$$

这个方程组从理论上看是简单的，而实用上未必行之有效. 这个方程组的所有解的集合组成特征空间.

1.4.3 定义 $A\in M_n$ 的相应于特征值λ 的特征空间的维数称为特征值λ 的几何重数. λ 作为特征多项式 $p_A(\cdot)$的零点的重数(至此，已经涉及了重数概念)称为特征值λ 的代数重数，一般说来，这两个概念是不同的. 在述及特征值时，如果不加限制地使用术语重数，那通常指的是代数重数. 我们将采用这个约定.

应该指出的是，几何重数正好是相应于某个特征值的线性无关特征向量的最大个数.

练习 证明，$A\in M_n$ 的一个特征值的几何重数不大于，且可能小于它的代数重数. 如果代数重数至少是 1，那么几何重数至少是 1. **提示**：假定λ 的几何重数是k，且设 $S\in M_n$ 是以 A 的相应于λ 的线性无关特征向量为其前k 个列的非奇异矩阵. 采用类似于(1.3.7)中所用过的证法，证明 $S^{-1}AS$ 有形式$\begin{bmatrix}\lambda I & * \\ 0 & *\end{bmatrix}$，$I\in M_k$，因而得出结论，$\lambda$ 的代数重数至少是k.

1.4.4 定义 如果矩阵 $A\in M_n$ 的某个特征值的几何重数严格小于其代数重数，就称 A 是亏损的. 如果每个特征值的几何重数都和其代数重数相同，就称 A 是非亏损的. 如果 $A\in M_n$ 的每个特征值恰好有几何重数 1(不考虑代数重数)，就称 A 是非减次的. 所有这些概念都是经典的，它们在某些场合被广泛采用.

我们指出，一个非减次的、非亏损的矩阵就是一个具有互不相同的特征值的矩阵. 另外，矩阵 $A\in M_n$ 可对角化，当且仅当 A 是非亏损的. 这只是重述(1.3.7)，它突出了每个特征值有足够多的相应线性无关特征向量的必要性.

1.4.5 例 尽管 A 和A^T 有相同的特征值，但是它们的相应于某个特征值的特征向量可能完全不同. 例如，设

58

$$A=\begin{bmatrix}2 & 3\\ 0 & 1\end{bmatrix}.$$

那么，A 的相应于特征值 2 的(一维)特征空间由$\begin{bmatrix}1\\0\end{bmatrix}$生成，而 A^T 的相应特征空间由$\begin{bmatrix}1\\3/2\end{bmatrix}$生成.

练习 验证(1.4.5)的细节.

很明显，迄今所阐述的特征值和征向量的理论，可以平行地对左乘以行向量来进行阐述. 诸特征值将会相同，但诸特征向量一般不同(即使考虑到行对应于列).

1.4.6 定义 非零向量 $y\in \mathbf{C}^n$ 称为 $A\in M_n$ 的相应于$\lambda\in\sigma(A)$左特征向量，是指

$$y^*A=\lambda y^*.$$

如果有必要明确，就称(1.1.3)中的向量为右特征向量，当上下文不要求区别时，就只说特征向量.

练习 证明，相应于 $A\in M_n$ 的特征值λ 的左特征向量 y 是 A^* 的相应于$\bar{\lambda}$ 的右特征向量.

同时 y 是 A^t 的相应于 λ 的右特征向量. 用例子说明，即使对于 $A\in M_n(\mathbf{R})$，左特征向量和右特征向量也未必相同.

从(0.6.2)可知，两个向量 x，$y\in\mathbf{C}^n$ 称为正交，是指 $y^*x=0$. 下面的结果称为双正交原理.

1.4.7 定理 如果 $A\in M_n$，且 λ，$\mu\in\sigma(A)$，$\lambda\neq\mu$，那么 A 的相应于 μ 的任一左特征向量与 A 的相应于 λ 的任一右特征向量正交.

证明： 设 $y\in\mathbf{C}^n$ 是 A 的相应于 μ 的左特征向量，而 $x\in\mathbf{C}^n$ 是 A 的相应于 λ 的右特征向量，用两种方式计算，y^*Ax：

$$y^*Ax=y^*(\lambda x)=\lambda(y^*x)$$
$$=(\mu y^*)x=\mu(y^*x).$$

59

因为 $\lambda\neq\mu$，所以 $\lambda y^*x=\mu y^*x$ 的唯一可能方式是 $y^*x=0$，即 x 与 y 正交. □

练习 如果 $A^*=A\in M_n$，即 A 是 Hermite 矩阵，且 A 有互不相同的特征值，证明存在 A 的 n 个两两正交的(右)特征向量. 从(1.1)节习题 8 可知，A 的特征值都是实数. **提示：** 因为 $A^*=A$，左特征向量与右特征向量相同. 应用(1.4.7).

在下一章将看到，在上述练习的陈述中，关于互不相同的特征值的假定是不必要的.

下面要指出，特征向量在相似下的变换方式是简单的，而特征值当然在相似下不变.

1.4.8 定理 设 A，$B\in M_n$. 如果 $x\in\mathbf{C}^n$ 是相应于 $\lambda\in\sigma(B)$的特征向量，且 B 经 S 与 A 相似，那么 Sx 是 A 的相应于特征值 λ 的特征向量.

证明： 如果 $B=S^{-1}AS$，且 $Bx=\lambda x$，那么 $S^{-1}ASx=\lambda x$，或 $ASx=\lambda Sx$. 因为 S 非奇异，且 $x\neq0$，所以 $Sx\neq0$，因而 Sx 是 A 的特征向量. □

练习 验证 $e=[1,1,1]^T$ 是

$$A=\begin{bmatrix}1&2&3\\3&2&1\\2&3&1\end{bmatrix}$$

的特征向量. 如果 $D=\mathrm{diag}(1,2,3)$，确定 $D^{-1}AD$ 的一个分量全为正的特征向量.

作为本节的最后一个结论，我们指出，可以利用特征向量得到有关主子矩阵的特征值的结果. 这个结果为一个特征值的几何重数与代数重数之间的不等式提供了又一个证明.

1.4.9 定理 设 $A\in M_n$ 且 $\lambda\in\mathbf{C}$ 已知，又设 $k\geqslant1$ 是某个正整数. 考虑下列三个命题：

(a) λ 是几何重数至少为 k 的 A 的特征值.

(b) 如果$\hat{A}\in M_m$ 是 A 的一个主子矩阵且 $m>n-k$，那么 λ 是$\hat{A}$的特征值.

(c) λ 是代数重数至少为 k 的 A 的特征值.

60

那么，(a) 蕴涵(b)，而(b)蕴涵(c). 特别是，特征值的代数重数至少等于它的几何重数.

证明： 假定(a)成立，且设$\hat{A}\in M_m$ 是 A 的阶数为 $m>n-k$ 的主子矩阵. 因为可以运用置换相似和(1.4.8)，所以不妨假定$\hat{A}$出现在 A 的左上角. 设 v_1，…，v_k 是 A 的相应于特征值 λ 的线性无关特征向量. 把 A 和每个 v_i 块分成

$$A=\begin{bmatrix}\hat{A} & * \\ * & *\end{bmatrix},\quad \hat{A}\in M_m;$$

$$v_i=\begin{bmatrix}u_i \\ w_i\end{bmatrix},\quad u_i\in \mathbf{C}^m,\quad w_i\in \mathbf{C}^{n-m},\quad i=1,2,\cdots,k.$$

因为向量 w_1，…，w_k 是维数为 $n-m<n-(n-k)=k$ 的空间中的 k 个向量，所以它们是相关的：因而存在不全为零的纯量 α_1，…，$\alpha_k\in\mathbf{C}$，使得 $\alpha_1w_1+\cdots+\alpha_kw_k=0$. 于是 $v\equiv\alpha_1v_1+\cdots+\alpha_kv_k=\begin{bmatrix}u \\ 0\end{bmatrix}\neq 0$，其中 $u=\alpha_1u_1+\cdots+\alpha_ku_k\neq 0$，且 $Av=\lambda v$. 把这个等式写成分块形式便得到

$$Av=\begin{bmatrix}\hat{A} & * \\ * & *\end{bmatrix}\begin{bmatrix}u \\ 0\end{bmatrix}=\begin{bmatrix}\hat{A}u \\ *\end{bmatrix}=\lambda v=\begin{bmatrix}\lambda u \\ 0\end{bmatrix}.$$

这说明 λ 是 $\hat{A}$ 的特征值. 这正是(b)中的结论.

现在假定(b)成立，并且想到恒等式(1.2.13)，这个恒等式把特征多项式 $p_A(t)$ 的导数与 A 的 n 个主子矩阵 A_1，…，A_n 的特征多项式 $p_{A_i}(t)$ 联系起来. 如果 $k=1$，那就没有什么可证的. 如果 $k>1$，那么(b)是说，λ 是每一个 A_i 的特征值，因而，$p_{A_i}(\lambda)=0$，且 $p'_A(\lambda)=0$. 如果 $k>2$，微分恒等式(1.2.13)得

$$p''_A(t)=\sum_{i=1}^{n}p'_{A_i}(t),\tag{1.4.10}$$

然后利用(1.2.13)，用每个 A_i 的诸主子矩阵的特征多项式之和代替等式右边的每个导数. 因为 A_i 的一个主子矩阵划去了一行和一列，所以它是 A 的阶数为 $n-2$ 的主子矩阵，把(b)中的假定和恒等式(1.2.13)用到每个 A_i 就得到 $p''_A(\lambda)=0$. 重复上述论证便可证明，对 $i=0$，1，…，$k-1$，各阶导数 $p_A^{(i)}(\lambda)$ 都为零，因而 λ 的代数重数至少为 k. □

习题

[61] 1. 证明，$A\in M_n$ 有秩1，当且仅当存在两个非零向量 x，$y\in\mathbf{C}^n$，使得

$$A=xy^*$$

并且证明：(a) 这个 A 至多有一个(代数重数是1的)非零特征值；(d) 这个特征值是 y^*x，(c) x 和 y 分别是相应于这个特征值的右特征向量和左特征向量. 特征值0的几何重数是多少？

2. 证明，秩为 k 的矩阵 $A\in M_n$ 可以写成

$$A=x^{(1)}y^{(1)*}+\cdots+x^{(k)}y^{(k)*},$$

其中，$x^{(i)}$，$y^{(i)}\in\mathbf{C}^n$，$i=1$，…，k，即 A 为 k 个秩1的矩阵之和. **提示**：求出 k 个线性无关的行和列，然后利用其余的行和列可以用它们来表示的事实.

3. 假定 $T\in M_n$ 是上三角矩阵，且它的互不相同的特征值 t_{11}，…，t_{nn} 沿对角线从左上角到右下角依次出现. 证明，存在 T 的相应于 t_{ii} 的一个右特征向量，它的最后 $n-i$ 个分量都是0，并且存在 T 的相应于 t_{ii} 的一个左特征向量，它的前 $i-1$ 个分量都是0. 如果 t_{ii} 不是互不相同的，那又会怎样呢？

4. 证明，在(1.2.7d)中所列出的矩阵的(仅有)特征值1有几何重数1. 描述它相应的特征空间.

5. 考虑分块三角矩阵

$$A=\begin{bmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix},\quad A_{ii}\in M_{n_i},\quad i=1,2.$$

证明 A 的各特征值是 A_{11} 的各特征值再加上 A_{22} 的各特征值(计相重特征值). 如果 $x\in\mathbf{C}^{n_1}$ 是 A_{11} 的相应于 $\lambda\in\sigma(A_{11})$ 的右特征向量，而 $y\in\mathbf{C}^{n_2}$ 是 A_{22} 的相应于 $\mu\in\sigma(A_{22})$ 的左特征向量，证明 $\begin{bmatrix}x\\0\end{bmatrix}\in\mathbf{C}^{n_1+n_2}$ 和 $\begin{bmatrix}0\\y\end{bmatrix}\in\mathbf{C}^{n_1+n_2}$ 是 A 的分别相应于 λ 的右特征向量和相应于 μ 的左特征向量. 关于 A 的分别相应于 λ 和 μ 的左特征向量和右特征向量，你能说些什么？你能把这些结果推广到具有任意多个对角子块的分块三角矩阵吗?

6. 如果 $A\in M_n$ 关于一个几何重数是 1 的特征值有正分量的左特征向量和右特征向量证明，除了与这些向量差一个倍数的向量以外，A 没有其他分量是非负的特征向量.

7. 在这个习题中，概述关于求 $A\in M_n$ 的最大特征值及其相应的特征向量的幂法. 我们作某些假定以简化论述，且使分析部分可以更为明确. 假定 $A\in M_n$ 有互不相同的特征值 $\lambda_1,\cdots,\lambda_n$，且恰有一个具有最大模 $\rho(A)$ 的特征值 λ_n. 如果 $x^{(0)}\in\mathbf{C}^n$ 与相应于 λ_n 的一个左特征向量不正交，证明序列 [62]

$$x^{(k+1)}=\frac{1}{(x^{(k)*}x^{(k)})^{1/2}}Ax^{(k)},k=0,1,2,\cdots$$

趋近于 A 的一个特征向量，而诸向量 $Ax^{(k)}$ 与 $x^{(k)}$ 中的某个给定分量的诸比值趋于 λ_n. **提示**：不失一般性，假定 $\lambda_n=1$，且设 $y^{(1)},\cdots,y^{(n)}$ 是相应于 $\lambda_1,\cdots,\lambda_n$ 的线性无关的特征向量. 向量 $x^{(0)}$ 可以唯一地表示成

$$x^{(0)}=\alpha_1y^{(1)}+\cdots+\alpha_ny^{(n)}$$

且 $\alpha_n\neq0$. 注意，$x^{(k)}=\alpha_1\lambda_1^ky^{(1)}+\cdots+\alpha_n\lambda_n^ky^{(n)}$，但差一个比例因子. 因为 $|\lambda_i|<1$，$|\lambda_i|^k\to0$，$i=1,\cdots,n-1$，从而这个和趋于 $y^{(n)}$ 的一个倍数.

8. 用幂法还可以计算其他的特征值(和特征向量)，不过需借助一个桥梁，称之为压缩，它给出一个阶数比 $A\in M_n$ 少 1 的方阵，其特征值是 A 余下来的特征值. 设 λ_n 和 $y^{(n)}$ 是(用幂法或其他方法算出的)A 的特征值和特征向量，且设 $S\in M_n$ 是第 1 列为 $y^{(n)}$ 的非奇异矩阵. 采用习题 7 中的记号，证明

$$S^{-1}AS=\begin{bmatrix}\lambda_n & *\\ 0 & A_1\end{bmatrix},$$

且 $A_1\in M_{n-1}$ 的特征值是 $\lambda_1,\cdots,\lambda_{n-1}$. 从 A_1 可以算出另外的特征值，然后重复施行压缩，如此等等.

9. 设 $A\in M_n$ 有特征值 $\lambda_1,\cdots,\lambda_{n-1},0$. 因而 $\operatorname{rank}A\leqslant n-1$. 再假定 A 的最后一行是其余各行的线性组合. (a)如果把 A 块分成

$$\begin{bmatrix}A_{11} & a_{12}\\ a_{21}^{T} & a_{22}\end{bmatrix},$$

其中 $A_{11}\in M_{n-1}$，证明，存在向量 $b\in\mathbf{C}^{n-1}$，使得

$$a_{21}^{T} = b^{T}A_{11} \text{ 和 } a_{22} - b^{T}a_{12}.$$

试用 A 的相应于 0 的左特征值来解释 b. (b)再证明 $A_{11}+a_{12}b^{T}\in M_{n-1}$ 有特征值 λ_1，…，λ_{n-1}. **提示**：考虑 A 的经

63

$$\begin{bmatrix} I & 0 \\ b^{T} & 1 \end{bmatrix}$$

的相似矩阵. 应指出的是，因为具有余下来的特征值的低阶矩阵是可以给出的，这便是压缩的另一种形式. 如果知道 A 的一个特征值 λ，那么，本习题所描述的过程可以应用于 $P(A-\lambda I)P^{-1}$，其中 P 是适当的置换矩阵.

10. 设 $T\in M_n$ 是非奇异矩阵，它的各列是 $A\in M_n$ 的左特征向量. 证明 $(T^{*})^{-1}$ 的各列是 A 的右特征向量.

64

第2章 酉等价和正规矩阵

下面，要研究一种特殊类型的相似，它跟矩阵分析的许多应用有密切关系.

2.0 导引

对一般的非奇异矩阵 $S\in M_n$，在第一章中初步研究了经 S 的相似问题. 对于某些很特殊的非奇异矩阵 S，称之为酉矩阵，它的逆有简单的形式：$S^{-1}=S^*$. $A\in M_n$ 经一个酉矩阵的相似，$A\to S^*AS$，不仅理论上比一般相似更简单(S^* 要比 S^{-1} 容易计算得多)，且有许多优点，这些优点通过以后的叙述会变得更加明显. 大致说来，酉相似优于一般的相似，因而，了解通过酉相似所获得的性质是很有用的. 但是，在酉相似下的等价类比在一般的相似下的等价类更细(两个矩阵可以相似，但不酉相似)，因而，相应地所能得到的每个等价类也较小. 在第3章中还要进一步研究一般的相似.

设 $A\in M_n$，假定 S 是非奇异矩阵，但不一定是酉矩阵，那么，变换 $A\to S^*AS$ 称为*相合，这将在第4章加以研究. 这个变换也是 M_n 上的一个等价关系，且有许多(与相似不同的)特点. 重要的是要意识到，通过酉矩阵的相似既是相似，也是*相合，并且是兼有相似性质与相合性质的最广泛的变换类.

2.1 酉矩阵

2.1.1 定义 我们知道，向量 x_1，…，$x_k\in \mathbf{C}^n$ 构成一个正交组，是指对所有的向量偶有，$x_i^*x_j=0$，$1\leqslant i<j\leqslant k$. 此外，如果诸向量还是正规化的，$x_i^*x_i=1$，$i=1$，…，$k$，那么该向量组称为标准正交组.

练习 如果$\{y_1$，…，$y_k\}$是由非零向量组成的正交组，证明由 $x_i=(y_i^*y_i)^{-1/2}y_i$，$i=1$，…，k 定义的向量组$\{x_1$，…，$x_k\}$是一个标准正交组.

2.1.2 定理 标准正交向量组线性无关.

证明： 假定：$\{x_1$，…，$x_k\}$是标准正交组，且 $0=\alpha_1x_1+\cdots+\alpha_kx_k$. 于是，因为诸向量 x_i 是正交的，所以 $0=0^*0=\sum_{i\neq j}\bar{\alpha}_i\alpha_jx_i^*x_j+\sum_{i=1}^{k}|\alpha_i|^2x_i^*x_i$；又诸向量 x_i 是正规化的，所有 $\sum_{i=1}^{k}|\alpha_i|^2x_i^*x_i=\sum_{i=1}^{k}|\alpha_i|^2=0$. 因此，所有 $\alpha_i=0$，故$\{x_1$，…，$x_k\}$是线性无关组.

练习 证明非零向量组成的正交组线性无关.

练习 证明，如果 x_1，…，$x_k\in\mathbf{C}^n$ 是正交组，那么，或者 $k\leqslant n$，或者诸向量 x_i 中至少有 $k-n$ 个等于零.

当然，一个无关组未必是正交组，不过，可以把 Gram-Schmidt 标准正交化过程(0.64)应用于该无关组，从而得到一个标准正交组，它也生成原向量组.

练习 证明任意 k 维实的或复的向量空间有一个标准正交基(一个由标准正交组组成的

基).

2.1.3 定义 设矩阵 $U\in M_n$，若 $U^*U=I$，就称 U 为酉矩阵. 若还有 $U\in M_n(\mathbf{R})$，就称 U 是实正交矩阵.

M_n 中的酉矩阵构成一个值得注意的重要集合. 在(2.1.4)中列出了关于 U 是酉矩阵的几个基本等价条件.

练习 若 $A\in M_n$ 是非奇异矩阵，且 $B\in M_n$ 使 $BA=I$. 证明，(a) B 是唯一的，(b) $AB=I$. 自然，记 $B=A^{-1}$. **提示**：非奇异性推出，对任意给定的 $y\in\mathbf{C}^n$，方程 $Ax=y$ 和 $x^TA=y^T$ 有唯一解. (逐列)证明 $AB_R=I$ 和(逐行)证明 $B_LA=I$ 有唯一解 B_L，$B_R\in M_n$，于是，用两种方式计算 B_LAB_R 便可证明 $B_L=B_R$.

66

2.1.4 定理 如果 $U\in M_n$，那么下列条件等价：

(a) U 是酉矩阵；

(b) U 是非奇异矩阵，且 $U^{-1}=U^*$；

(c) $UU^*=I$；

(d) U^* 是酉矩阵；

(e) U 的各列组成一个标准正交组；

(f) U 的各行组成一个标准正交组；

(g) 如果 $x\in\mathbf{C}^n$ 且 $y=Ux$，那么，y 的 Euclid 长度与 x 的相同：即 $y^*y=x^*x$.

证明：条件(a)蕴涵条件(b)，这是因为，U^{-1}是唯一的(只要它存在)，且 U 左乘以它就得 I；酉矩阵的定义保证 U^* 就是这样的矩阵. 因为 $BA=I$，当且仅当 $AB=I$(对于 A，$B\in M_n$)，所以(b)蕴涵(c). 因为$(U^*)^*=U$，(c)推出 U^* 适合酉矩阵所必需的条件；即(c)蕴涵(d). 由于上述每个推理都可类似地反推回去，所以(a)～(d)是等价的.

根据矩阵乘法的技巧，likewise设 $u^{(i)}$ 表示 U 的第 i 列，$i=1$，…，n. 条件 $U^*U=I$ 的意思是说

$$u^{(i)*}u^{(j)}=\begin{cases}0,\text{如果 } j\neq i,\\1,\text{如果 } j=i.\end{cases}$$

因此，$U^*U=I$ 的另一种解释是，U 的各列是标准正交组，因而(a)与(e)等价. 同理(d)与(f)等价.

如果(a)成立，且 $y=Ux$，则 $y^*y=x^*U^*Ux=x^*Ix=x^*x$，因而(a)蕴涵(g). 另一方面，为了验证逆命题成立，需要作稍微复杂的计算. 不过，本书后面要给出的方法将使这种计算更为直接. 首先考虑 $n=2$ 的情形，假定(g)成立，并且设 $x=\begin{bmatrix}1\\0\end{bmatrix}$，我们得知，$1=x^*x=y^*y=x^*U^*Ux=U^*U$ 的1，1元. 类似地，设 $x=\begin{bmatrix}0\\1\end{bmatrix}$，得出 U^*U 的 2，2 元也是 1. 因而 U^*U 必须有形状

$$\begin{bmatrix}1 & a\\ \bar{a} & 1\end{bmatrix},$$

其中，a 是 U 的第 1 列与第 2 列的内积，而 $\bar{a}$ 是第 2 列与第 1 列的内积. 在(g)中设 $x=\begin{bmatrix}1\\1\end{bmatrix}$，然后再作计算，因而得知，$2=x^*x=y^*y=x^*U^*Ux=2+(a+\bar{a})$. 设 $x=\begin{bmatrix}1\\i\end{bmatrix}$，得知 $2=2+i(a-\bar{a})$. 因此，$a+\bar{a}=2\mathrm{Re}\,a=0$，且 $a-\bar{a}=2i\mathrm{Im}\,a=0$，从而 $a=0$. 这就是说，如果对所有 $x\in\mathbf{C}^2$，$x^*U^*Ux=x^*x$，那么 $U^*U=I$；即 U 是酉矩阵(如果 $U\in M_2$). 现在考虑 $n>2$ 的情形，且设 $A=U^*U$. 假定 $x\in\mathbf{C}^n$ 是这样一个向量. 除了第 i 个和第 j 个分量以外，$i\leqslant j$，它的所有分量都是 0. 于是

67

$$x^*Ax=[\bar{x}_i,\bar{x}_j]A(\{i,j\})\begin{bmatrix}x_i\\x_j\end{bmatrix},$$

[见(0.7.1)关于子矩阵的记号]，因此，刚才证明了(g)蕴涵 $A(\{i,j\})=I\in M_2$. 因为 i 和 j 的任意性，得知 A 的每个 2×2 主子矩阵都是 2×2 单位矩阵. 正是这个 A 有 $A=I\in M_n$，又因为 $n=1$ 的情形是显然的，因此得出(g)蕴涵(a)，证毕. □

2.1.5 定义 一个线性变换 T：$\mathbf{C}^n\to\mathbf{C}^m$ 称为 Euclid 等距变换，是指对所有 $x\in\mathbf{C}^n$，有 $x^*x=(Tx)^*(Tx)$. 定理(2.1.4)是说，一个复方阵 $U\in M_n$ 是一个(经由 U：$x\to Ux$ 的)Euclid 等距变换当且仅当 U 是酉矩阵. 关于其他形式的等距变换见 5.2 节.

练习 设

$$T(\theta)=\begin{bmatrix}\cos\theta & \sin\theta\\-\sin\theta & \cos\theta\end{bmatrix},$$

其中 θ 是一个实参数. (a)如果 $U\in M_2(\mathbf{R})$，证明，U 是实正交矩阵，当且仅当 $U=T(\theta)$ 或

$$U=\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}T(\theta)$$

对某个 $\theta\in\mathbf{R}$ 成立. (b)如果 $U\in M_2(\mathbf{R})$，证明，U 是实正交矩阵，当且仅当 $U=T(\theta)$ 或

$$U=\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}T(\theta)$$

对某个 $\theta\in\mathbf{R}$ 成立. 我们用一个参数 θ 给出了 2×2 实正交矩阵的两种不同的表示形式. 试从几何上解释它们.

2.1.6 论断 如果 U，$V\in M_n$ 是酉(相应地，实正交)矩阵，那么，乘积 UV 也是酉(相应地，实正交)矩阵.

练习 利用(2.1.4)的(b)证明(2.1.6).

68

练习 如果 $\{x_1, x_2, \cdots, x_k\}\subseteq\mathbf{C}^n$ 是标准正交组，且 $U\in M_n$ 是酉矩阵，证明 $\{Ux_1, \cdots, Ux_k\}$ 是标准正交组.

2.1.7 论断 M_n 中的酉(相应地，实正交)矩阵的集合构成一个群. 一般称这个群为 $n\times n$ 酉(相应地，正交)群. 它是 $GL(n, \mathbf{C})$ 的一个子群[见(0.5)节].

练习 我们知道，一个群是这样一个集合，它在一个结合的二元运算("乘法")下是封闭的，且使得关于这个运算的单位元和逆仍在该集合中验. 证(2.1.7). **提示**：用(2.1.6)验证封

闭性：矩阵乘法是结合的；$I\in M_n$ 是酉矩阵；且 $U^*=U^{-1}$ 也是酉矩阵.

M_n 中的酉矩阵的集合(群)有另外一些很重要的性质. 一个矩阵序列的"收敛性"概念和"极限"的概念将在第 5 章明确地给出，不过，这里可以从每个 i，j 元的"收敛性"和极限的观点来理解这些概念. 定义恒等式 $U^*U=I$ 意味着 U 的每个列有 Euclid 长度 1，因而 $U=[U_{ij}]$没有一个元 U_{ij} 的绝对值比 1 大. 如果把酉矩阵的集合看作 $\mathbf{C}^{n^2}$ 的一个子集，这说明它是一个有界子集. 如果 $U_k=[u_{ij}^{(k)}]$是一个酉矩阵序列，$k=1, 2, \cdots$，使得所有 $i, j=1, 2, \cdots, n$，$\lim\limits_{k\to\infty}u_{ij}^{(k)}\equiv u_{ij}^{(0)}$ 存在，那么由恒等式 $U_k^*U_k=I$，$k=1, 2, \cdots$. 我们看到$\lim\limits_{k\to\infty}U_k^*U_k=U_0^*U_0=I$，其中 $U_0=[u_{ij}^{(0)}]$. 因此，极限矩阵 U_0 也是酉矩阵. 这就是说，酉矩阵的集合是 $\mathbf{C}^{n^2}$ 的一个闭子集.

由于有限维 Euclid 空间的一个有界闭子集是紧集(见附录 E)，由此可知，M_n 中的酉矩阵集合(群)是紧的. 眼下，这个结论的最重要的推论是下述关于酉矩阵的选择原理.

2.1.8 引理 设 U_1，U_2，…是 M_n 中的已知的酉矩阵序列. 则存在一个子序列 U_{k_1}，U_{k_2}，…使得当 $i\to\infty$时，U_{k_i} 的每个元(作为复数列)收敛于一个酉矩阵 U_0 的对应元.

证明： 这里只要求我们，总可以从一个紧集的任一无限序列中选出一个收敛子序列. 我们已经知道，如果一个酉矩阵序列收敛于某个矩阵，那么其极限矩阵一定是酉矩阵. [69] □

由引理所确定的酉极限不一定是唯一的；它可能与子序列的选择有关.

练习 考虑酉矩阵序列

$$U_k=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}^k,\quad k=1,2,\cdots.$$

证明它可以有两个子序列极限.

练习 选择原理(2.1.8)也适应于正交群；即，一个实正交矩阵序列有一个收敛于实正交矩阵的子序列. 在实的情形，试通过复述同样的推理验证这个结论.

酉群的紧性在本书的其他地方要用到它.

一个酉矩阵有 U^{-1}等于U^* 的性质. 一种推广酉矩阵的概念的方式是要求 U^{-1}相似于 U^*. 容易看出，对所有非奇异矩阵 $A\in M_n$，这些矩阵的集合可以刻划为映射 $A\to A^{-1}A^*$ 的值域.

2.1.9 定理 设 $A\in M_n$ 是非奇异矩阵，那么，A^{-1}相似于 A^*，当且仅当存在非奇异矩阵 $B\in M_n$，使得 $A=B^{-1}B^*$.

证明： 对于某个非奇异矩阵 $B\in M_n$，如果有 $A=B^{-1}B^*$，那么，$A^{-1}=(B^*)^{-1}B$，并且 $B^*A^{-1}(B^*)^{-1}=B(B^*)^{-1}=(B^{-1}B^*)^*=A^*$，这样，$A^{-1}$可经相似矩阵 B^* 相似于 A^*. 反过来，如果 A^{-1}相似于 A^*，那么，存在非奇异矩阵 $S\in M_n$，使得 $SA^{-1}S^{-1}=A^*$. 对 $\theta\in R$，令 $S_\theta=e^{i\theta}S$，且注意到 $S_\theta A^{-1}S_\theta^{-1}=e^{i\theta}SA^{-1}(e^{-i\theta}S^{-1})=SA^{-1}S^{-1}=A^*$. 另一方面，$S_\theta=A^*S_\theta A$，且 $S_\theta^*=A^*S_\theta^*A$. 相加这两个恒等式便得到 $H_\theta=A^*H_\theta A$，其中，$H_\theta\equiv S_\theta+S_\theta^*$ 是 Hermite 矩阵. 如果 H_θ 是奇异矩阵，那么，存在某个非零 $x\in\mathbf{C}^n$，使得 $0=H_\theta x=S_\theta x+S_\theta^*x$，因而，$-x=S_\theta^{-1}S_\theta^*x=e^{-2i\theta}S^{-1}S^*x$，且$S^{-1}S^*x=-e^{2i\theta}x$. 选取值 $\theta=\theta_0\in[0, 2\pi]$，使$-e^{2i\theta_0}$不是 $S^{-1}S^*$ 的特征值；所得到的 Hermite 矩阵 $H=H_{\theta_0}$是非奇异矩阵，且有性质 $H=A^*HA$.

现在，选取任一复数 α，使 $|\alpha|=1$，且 α 不是 A^* 的特征值. 令 $B\equiv\beta(\alpha I-A^*)H$，其中

复参数 $\beta\neq 0$ 是有待选定的，不难看出，B 是非奇异矩阵．为了使 $A=B^{-1}B^*$ 或 $BA=B^*$，经计算，$B^*=H(\bar{\beta}\bar{\alpha}I-\bar{\beta}A)$，而 $BA=\beta(\alpha I-A^*)HA=\beta(\alpha HA-A^*HA)=\beta(\alpha HA-H)=H(\alpha\beta A-\beta I)$．如果能选定一个非零 β 使 $\beta=-\bar{\beta}\bar{\alpha}$，就完成了证明，而当 $\alpha=e^{i\theta}$ 时，只要取 $\beta=e^{i(\pi-\theta)/2}$就行了． □ 70

习题

1．如果 $U\in M_n$ 是酉矩阵，证明 $|\det U|=1$．

2．如果 $\lambda\in\sigma(U)$，而 $U\in M_n$ 是酉矩阵，证明 $|\lambda|=1$．**提示**：利用等矩性质(2.1.4g)．

3．给定实参数 θ_1，θ_2，…，θ_n，证明

$$U=\mathrm{diag}(e^{i\theta_1},e^{i\theta_2},\cdots,e^{i\theta_n})$$

是酉矩阵．

4．说明实对角正交矩阵的特征．

5．证明 M_n 中的置换矩阵(0.9.5)是正交矩阵，并证明置换矩阵构成实正交矩阵群的一个子群(这个子集本身是一个群)．在 M_n 中有多少不同的置换矩阵？

6．你能给出 3×3 正交群的一种参数表示形式吗？想一想本节中给出的 2×2 正交群的两种表示形式．

7．对定理(2.1.4)中条件(g)蕴涵条件(a)的下述另一个证明作详细的论述．证明(g)表示对所有 $x\in\mathbf{C}^n$ 有 $x^*(U^*U-I)x=0$．设 $H\equiv U^*U-I$ 并且注意到 $H=H^*$．考虑 $0=(x+e^{i\theta}y)^*H(x+e^{i\theta}y)$，对所有 x，$y\in\mathbf{C}^n$ 和所有 $\theta\in\mathbf{R}$．展开这个等式并证明对所有 x，$y\in\mathbf{C}^n$ 有 $x^*Hy=0$．系统地选取 x 和 y 可得出 $H=0$．

8．适合 $AA^T=I$ 的矩阵 $A\in M_n$ 称为正交矩阵．一个实正交矩阵是酉矩阵，但一个非实正交矩阵不一定是酉矩阵．(a)设

$$K=\begin{bmatrix}0&1\\-1&0\end{bmatrix}\in M_2(\mathbf{R}).$$

证明，对所有 $t\in\mathbf{R}$，$A(t)=(\cosh t)I+(i\sinh t)K\in M_2$ 是正交矩阵，但只有 $t=0$ 时 $A(t)$ 才是酉矩阵．其中，双曲函数定义为 $\cosh t=(e^t+e^{-t})/2$，$\sinh t=(e^t-e^{-t})/2$．(b)证明，与酉矩阵不同的是，复正交矩阵的集合不是有界集，因而它不是紧集．(c)证明，跟酉矩阵一样，一个给定阶数的所有复正交矩阵的集合构成一个群．尽管如此，通常的做法是，*正交群*这一术语专指一个由给定阶数的所有实正交矩阵组成的较小的(紧)群．(d)若 $A\in M_n$ 是正交矩阵，证明，$|\det A|=1$，但可能有特征值 λ 适合 $|\lambda|\neq 1$．**提示**：考虑(a)中的 $A(t)$，证明 $|\lambda(t)|$ 可以任意大．(e)若 $A\in M_n$ 是正交矩阵，证明 $\bar{A}$，A^T 和 A^* 都是正交矩阵且 A 是非奇异的．A 的诸行或诸列构成正交组吗？(f)说明对角正交矩阵的特征．与习题 4 比较．为避免混淆，有的作者称不一定是实矩阵的正交矩阵为*复正交矩阵*，但是在文献中一般不明确这种区别；术语*正交矩阵*有时是指我们所称的*实正交矩阵*． 71

9．如果 $U\in M_n$ 是酉矩阵，证明 $\bar{U}$，U^T 和 U^* 都是酉矩阵．

10．如果 $U\in M_n$ 是酉矩阵，证明，x，$y\in\mathbf{C}^n$ 正交，当且仅当 Ux 与 Uy 正交．

11．称非奇异矩阵 $A\in M_n$ 是*斜正交矩阵*，是指 $A^{-1}=-A^T$．证明，A 是斜正交矩阵，当

且仅当 $\pm iA$ 是正交矩阵. 更一般地，如果 $\theta\in\mathbf{R}$，证明，$A^{-1}=e^{i\theta}A^T$，当且仅当 $e^{i\theta/2}A$ 是正交矩阵. 当 $\theta=\pi$ 时，它是什么？如果 $\theta=0$ 呢？

12. 证明，如果 $A\in M_n$ 相似于一个酉矩阵，那么 A^{-1} 相似于 A^*.

13. 考虑矩阵 $A=\mathrm{diag}\left(2,\ \frac{1}{2}\right)\in M_2$. 证明相似于酉矩阵的所有矩阵的集合是其逆 A^{-1} 相似于 A^* 的所有矩阵 A 的集合的一个真子集.

14. 证明 M_n 中的酉矩阵群与 M_n 中的复正交矩阵群之交是 M_n 的实正交群. **提示：** 考虑 $U=A+iB$，其中 U，A，$B\in M_n$，且 A、B 是实矩阵. 如果 U 既是酉矩阵，又是复正交矩阵，证明 $B^TB=0$，从而，对每个标准基向量 $e_i\in\mathbf{R}^n$，$(Be_i)^T(Be_i)=0$，因此 B 的每个列都是 0.

进一步阅读 想了解关于适合定理(2.1.9)的条件的推广的酉矩阵的更多信息，可参看 C. R. DePrima and C. R. Johnson, "The Range of $A^{-1}A^*$ in $GL(n, \mathbf{C})$," *Linear Algebra Appl.* 9(1974), 209-222.

2.2 酉等价

因为酉矩阵 U 有 $U^*=U^{-1}$，所以，如果 U 是酉矩阵，那么由 $A\to U^*AU$ 给出的 M_n 上的变换是相似变换. 这种特殊的相似形式称为酉相似或酉等价.

2.2.1 定义 设矩阵 A，$B\in M_n$，如果存在酉矩阵 $U\in M_n$，使得 $B=U^*AU$，就称 B 酉等价于 A. 如果 U 可以取实矩阵(因而是实正交矩阵)，那么就称 B(实)正交等价于 A. [72]

练习 证明酉等价是一个等价关系.

2.2.2 定理 如果 $A=[a_{ij}]$ 和 $B=[b_{ij}]\in M_n$ 酉等价，那么

$$\sum_{i,j=1}^{n}|b_{ij}|^2=\sum_{i,j=1}^{n}|a_{ij}|^2.$$

证明： 作矩阵乘法后便可看出 $\sum_{i,j}|a_{ij}|^2=\mathrm{tr}\,A^*A$. 因此只要验证 $\mathrm{tr}\,B^*B=\mathrm{tr}\,A^*A$ 就够了. 如果 $B=U^*AU$，则因为迹是相似不变量，就有 $\mathrm{tr}\,B^*B=\mathrm{tr}\,U^*A^*UU^*AU=\mathrm{tr}\,U^*A^*AU=\mathrm{tr}\,A^*A$. □

练习 定理(2.2.2)说明 $\mathrm{tr}\,A^*A$ 是酉相似不变量. 不用 A^*A 还可以作出另一个证明. 不过要利用上一节的结果：用一个酉矩阵乘一个向量不会改变其 Euclid 长度，要注意的是，从左做起的矩阵乘法是乘被乘矩阵的各列，而从右做起的矩阵乘法是乘被乘矩阵的各行.

练习 证明

$$\begin{bmatrix}3 & 1\\ -2 & 0\end{bmatrix}\quad 和\quad \begin{bmatrix}1 & 1\\ 0 & 2\end{bmatrix}$$

相似，但不酉等价.

因为酉等价蕴涵相似，但是反之不成立，所以，与相似等价关系相比，酉等价关系把 M_n 划分成更细的等价类. 跟相似一样，酉等价对应一个基变换，不过它是一种特殊形式的基变换——它把一个标准正交基变成另一个标准正交基. 基的标准正交变换不改变向量的各分量的绝对值的平方和，而这个量在基的一个非标准正交变换下是可以改变的. 因为求共轭转置要比

计算逆容易得多，所以在计算上，酉等价关系要比相似关系更简单．在出现舍入误差的情况下，它也能较好地保证精确性．因此，在数值计算中，它是可取的．关于这方面的细致理由就不在这里阐述了，但是，可以作一个直观的解释，那就是用一个酉矩阵做乘法有保持长度的性质．

下述两种特殊的(又很简单的)酉矩阵形式给出两种酉等价变换，它们在特征值的计算中非常重要． [73]

2.2.3 例：平面旋转 设

$$
U(\theta;i,j)=\begin{bmatrix}
1 & & & & & & & & \\
 & \ddots & & & & & & & \\
 & & 1 & & & & & & \\
 & & & \cos\theta & 0 & \cdots & 0 & -\sin\theta & & \\
 & & & 0 & 1 & & & & & \\
 & & & \vdots & & \ddots & & & & \\
 & & & 0 & 0 & & 1 & & & \\
 & & & \sin\theta & 0 & \cdots & 0 & \cos\theta & & \\
 & & & & & & & & 1 & \\
 & & & & & & & & & \ddots \\
 & & & & & & & & & & 1
\end{bmatrix}
\begin{matrix} \\ \\ \\ i\text{ 行} \\ \\ \\ \\ j\text{ 行} \\ \\ \\ \\ \end{matrix}
$$

i 列　　　j 列

它相当于一个单位矩阵，只是它的 i，i 元和 j，j 元代之以 $\cos\theta$，而 i，j 元(相应地 j，i 元)用 $-\sin\theta$(相应地 $\sin\theta$)来代替．

练习 对于任一对指标 $1\leqslant i<j\leqslant n$ 和任一角参数 $0\leqslant\theta<2\pi$，验证 $U(\theta;\ i,\ j)$ 是 $M_n(\mathbf{R})$ 中的正交矩阵．矩阵 $U(\theta;\ i,\ j)$ 只是在 i，j 坐标平面中作一个(θ 角的)旋转．值得指出的是，左乘以 $U(\theta;\ i,\ j)$ 仅对被乘矩阵的 i 行和 j 行有影响，而右乘以 $U(\theta;\ i,\ j)$ 仅对 i 列和 j 列有影响，因此，经 $U(\theta;\ i,\ j)$ 的酉等价只使 i，j 行和 i，j 列有变化．经平面旋转的酉等价是计算特征值的 Jacobi 法和 Givens 法的基本特征．

2.2.4 例 Householder 变换．设 $w\in\mathbf{C}^n$ 是非零向量，定义 $U_w\in M_n$ 为 $U_w=I-tww^*$，其中 $t=2(w^*w)^{-1}$．注意，$ww^*\in M_n$，而 w^*w 是一个正纯量．如果 w 已正规化($w^*w=1$)，那么，t 就是 2，而 U_w 就是 $I-2ww^*$．我们常常是在假定 w 已经正规化的前提下给出矩阵 U_w．

练习 证明，U_w 在补子空间 $w^\perp$ 上相当于一个单位变换，而在由 w 生成的一维子空间上相当于一个反射；即，如果 $x\perp w$，则 $U_w x=x$，而 $U_w W=-W$． [74]

练习 证明，U_w 既是酉矩阵，也是 Hermite 矩阵($U_w^*=U_w$)．形如 U_w 的矩阵称为 Householder 变换．经 U_w 的酉等价有时也叫做 Householder 变换．这些变换在本书中要多次出现，其中包括计算特征值的 Householder 法(见习题 4)以及其他酉简化方法．值得指出的是，Householder 变换一般变更它们所作用的矩阵或向量的所有元素，但是，它们为许多应用提供

非常有效而精确的简化.

定理(2.2.2)给出了两个已知矩阵酉等价的一个必要而不充分的条件. 我们能在此基础上再添加更多的恒等式，这些恒等式共同给出两个已知矩阵酉等价的必要充分条件. 其中，下述简单概念起着关键的作用. 设 s，t 是两个给定的非交换变元，称 s，t 的非负整数幂的任一有限形式乘积

$$W(s,t) = s^{m_1}t^{n_1}s^{m_2}t^{n_2}\cdots s^{m_k}t^{n_k}, \quad m_1,n_1,\cdots,m_k,n_k \geqslant 0 \tag{2.2.5}$$

为关于 s 和 t 的字. 字 $W(s,t)$的次数是非负整数 $m_1+n_1+m_2+n_2+\cdots+m_k+n_k$，即字中的所有指数之和. 如果已知 $A\in M_n$，我们可以形式地定义 A 和 A^* 的字为

$$W(A,A^*) = A^{m_1}(A^*)^{n_1}A^{m_2}(A^*)^{n_2}\cdots A^{m_k}(A^*)^{n_k}.$$

由于 A 与 A^* 的幂未必交换，试图重排乘积中的项来简化 $W(A,A^*)$不一定可行.

如果 A 酉等价于某个 $B\in M_n$，那么，对某个酉矩阵 $U\in M_n$，$A=UBU^*$，且不难算出

$$\begin{aligned}W(A,A^*) &= (UBU^*)^{m_1}(UB^*U^*)^{n_1}\cdots(UBU^*)^{m_k}(UB^*U^*)^{n_k}\\&= UB^{m_1}U^*U(B^*)^{n_1}U^*\cdots UB^{m_k}U^*U(B^*)^{n_k}U^*\\&= UB^{m_1}(B^*)^{n_1}\cdots B^{m_k}(B^*)^{n_k}U^*\\&= UW(B,B^*)U^*.\end{aligned}$$

因而，$\operatorname{tr} W(A,A^*)=\operatorname{tr} UW(B,B^*)U^*=\operatorname{tr}W(B,B^*)$如果取字 $W(s,t)=ts$，就得到定理(2.2.2)中的恒等式.

对所有可能的字 $W(s,t)$，上述结果给出了两个矩阵是酉等价的无限多个必要条件. 下面，我们不加证明地叙述 W·Specht 的一个定理，这个定理确认这无限多个必要条件同时也是充分条件.

75

2.2.6 定理 两个已知矩阵 A、$B\in M_n$ 酉等价，当且仅当

$$\operatorname{tr} W(A,A^*) = \operatorname{tr} W(B,B^*) \tag{2.2.7}$$

对于两个非交换元的每个字 $W(s,t)$都成立.

Specht 定理只能用来证明两个已知矩阵不酉等价. 然而，除非在一些特殊情形(见习题6)，它在证明两个已知矩阵是酉等价时可能失效，因为必须验证无限多个条件. 幸运的是，有一个定理对 Specht 定理作了改进，它属于 C. Pearcy. 这个定理说明，只要对有限多个字验证迹恒等式(2.2.7)就够了.

2.2.8 定理 两个已知矩阵 A，$B\in M_n$ 酉等价，当且仅当 $\operatorname{tr} W(A,A^*)=\operatorname{tr} W(B,B^*)$对次数至多是 $2n^2$ 的每个字 $W(s,t)$都成立.

Pearcy 定理中的有限限制是对 Specht 定理的一大改进，但是，如我们所知，这还是非常保守的. 当 $n=2$ 时，只要对三个字 $W(s,t)=s$，s^2 和 ts 验证迹恒等式就够了. 无需考虑次数至多是 $2(2)^2=8$ 的所有字. 当 $n=3$ 时，只要对 9 个字 $W(s,t)=s$，s^2，ts，s^3，ts^2，t^2s^2，$tsts$，ts^2ts 和 ts^2t^2s 验证迹恒等式而用不着考察次数至多是 $2(3)^2=18$ 的所有字.

习题

1. 设 $A=[a_{ij}]\in M_n(\mathbf{R})$ 是对称矩阵($A^T=A$)，但非对角矩阵，并假定选下标 $i\neq j$ 使得 $|a_{ij}|$ 越大越好. 用$(a_{ii}-a_{jj})/2a_{ij}=\cot(2\theta)$定义 θ，且设 $U(\theta;\ i,\ j)$是如(2.2.3)中所得到的平面旋转. 用(2.2.2)证明，如果 $B=U(\theta;\ i,\ j)^* AU(\theta;\ i,\ j)=(b_{ij})$，那么 $\sum_{i\neq j}|b_{ij}|^2<\sum_{i\neq j}|a_{ij}|^2$. 说明累次应用这样的平面旋转(用上述相同的方式对 B 和它后面的矩阵作相应的选择)会缩小各非对角元的平方和而保持所有元的平方和不变；每做一步，所得到的矩阵就比前一步更接近于对角矩阵. 这就是计算一个实对称矩阵的特征值的 Jacobi 法. 这个方法产生一个实对角矩阵的矩阵序列. 为什么极限矩阵的诸对角元必定是 A 的诸特征值?

2. 关于实对称矩阵(或一般的实矩阵)特征值的 Givens 法也是利用平面旋转，但采用的方式有所不同. 证明，一个对称矩阵 $A=[a_{ij}]\in M_n(\mathbf{R})$经若干平面旋转后正交等价于一个三对角(对称)矩阵，而一般的 $A\in M_n(\mathbf{R})$经若干平面旋转正交等价于一个(下)Hessenberg 矩阵. 关于三对角矩阵和 Hessenberg 矩阵可见(0.9.9)和(0.9.10). **提示**：选一个平面旋转 $U_{1,3}$使得 $U_{1,3}^*AU_{1,3}$的 1，3 元为 0. 选另一个平面旋转使 1，4 元为 0，继续做下去直到 1，n 元. 然后转到 2，4 元，接着做下去. 注意，这样一步步做下去不会影响前面已经选定的 0 元；并且正交等价保持对称性不变. 一个三对角矩阵的特征多项式是容易确定的，而其特征值可用求根的方法来确定. 注意到经过有限多次平面旋转后 Givens 法得到一个三对角矩阵，但它不会显现特征值或特征向量，不过利用另外的计算一定能得到它们. Jacobi 法一般不会经有限多次平面旋转后中止，但它的目标是产生一个对角矩阵，而且得到一个标准正交特征向量组.

76

3. 证明，每个矩阵 $A\in M_n$ 酉等价于一个有相同主对角元的矩阵. **提示**：(a)若 $A\in M_2$，考察 $A-(1/2)(\operatorname{tr}A)I$，证明只需考虑 $\operatorname{tr}A=0$ 的情形就行了. 若 $x\in\mathbf{C}^2$ 是一个使 $x^*Ax=0$ 的单位向量，设 $U=[x,\ y]\in M_2$ 是一个酉矩阵，说明 U^*AU 的 1，1 元为零，并且利用迹条件证明其 2，2 元也为零. 为了求这个向量 x，设 w、z 是属于 A 的两个特征值$\pm\lambda$ 的特征向量. 如果 $\lambda=0$，取 $x=w$. 如果$\lambda\neq0$，设 $x(\theta)=e^{i\theta}w+z$；证明，对所有 $\theta\in\mathbf{R}$ 有 $x(\theta)\neq0$，而对某个 $\theta\in\mathbf{R}$ 有 $x(\theta)^*Ax(\theta)=0$；对这个 θ，设 $x=x(\theta)/[x(\theta)^*x(\theta)]^{1/2}$. 如果 $A\in M_2(\mathbf{R})$，容易构造一个(实)平面旋转 $U=U(\theta;\ 1,\ 2)$使 U^TAU 的对角元相等，但这在复的情形是行不通的.

定义$f(A)=\max\{|a_{ii}-a_{jj}|:\ i,\ j=1,\ 2,\ \cdots,\ n\}$，并且设 $A_2=\begin{bmatrix}a_{ii} & a_{ij}\\ a_{ji} & a_{jj}\end{bmatrix}$，其中指标偶 i，j 适合 $f(A)=|a_{ii}-a_{jj}|$. 设 $U_2\in M_2$ 是使$U_2^*A_2U_2$ 有相同主对角元的酉矩阵. 如同(2.2.3)中从2×2 平面旋转构造 $U(\theta;\ i,\ j)$，用同样的方式从 U_2 构造 $U(i,\ j)\in M_n$，并且证明，如果极大对角元之差不是多个可取差，则 $f(U(i,\ j)^*AU(i,\ j))<f(A)$；否则可以重复上述构造. 可以断定，如果 $f(A)\neq0$，则有酉矩阵 $U\in M_n$ 使 $f(U^*AU)<f(A)$. 设 $R(A)=\{U^*AU:\ U\in M_n$ 是酉矩阵$\}$，证明 $R(A)$是紧集并注意 f 是 $R(A)$上的连续函数. 设 $C\in R(A)$适合 $f(C)=\min\{f(B):\ B\in R(A)\}$，证明 $f(C)$不成立，因而从 $f(C)=0$ 得出结论成立.

4. 证明，用 Householder 变换可以把一个具有 Euclid 长度 $r=(x^Tx)^{1/2}$的任一向量 $x\in\mathbf{R}$ 变成长度为 r 的任一其他向量 $y\in\mathbf{R}^n$，$y\neq x$. **提示**：试用 U_w 证明，其中 $w=x-y$. 如果 x，$y\in\mathbf{C}^n$ 你能得出什么结论?

77

5. 计算 $A\in M_n(\mathbf{R})$ 的特征值的 Householder 法如同 Givens 法，先把 A 化简成(上) Hessenberg 矩阵(或在对称的情形化为三对角矩阵). 构造性地证明，其第 j 列中的第$(j+1)$个元素以后均为 $0(j=1, \cdots, k)$的形如

$$\left[\begin{array}{cccccc:c} * & * & * & \cdots & * & & \\ * & * & * & & & & \\ & * & * & & \vdots & & \\ & & * & \ddots & & & * \\ & 0 & & \ddots & * & & \\ & & & & * & & \\ \hdashline & & 0 & & & & * \end{array}\right] \begin{array}{l} \\ \\ \\ \\ \leftarrow k\text{ 行} \\ \leftarrow (k+1)\text{ 行} \\ \\ \end{array}$$

$\uparrow$
k 列

的矩阵，经一个用 Householder 变换所确定的实正交相似可以变成一个同形的矩阵. 由此得出，任一个矩阵 $A\in M_n(\mathbf{R})$都可以经一系列($n-2$ 个)Householder 相似化简成上 Hessenberg 矩阵，而一个对称矩阵 $A\in M_n(\mathbf{R})$可用同样的方式化简成三对角矩阵. 提示：对于第$(k+1)$列，选一 Householder 变换 $U\in M_{n-k}$把出现在对角线下方的诸元素组成的$(n-k)$维向量变成$[1, 0, \cdots, 0]^T\in\mathbf{R}^{n-k}$的一个适当倍数. 于是经由正交矩阵

$$\begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix}\in M_n$$

确定的相似可变换整个矩阵，并且看到所想要的 0 子块出现了.

6. 设 $A\in M_n$ 和 B, $C\in M_n$ 是给定的. 用 Specht 定理(2.26)或 Pearcy 定理(2.2.8)证明 B 与 C 酉等价，当且仅当下列诸条件下的任一个成立：

(a) $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$和$\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$酉等价.

(b) $\begin{bmatrix} B & & & 0 \\ & B & & \\ & & \ddots & \\ 0 & & & B \end{bmatrix}$和$\begin{bmatrix} C & & & 0 \\ & C & & \\ & & \ddots & \\ 0 & & & C \end{bmatrix}$酉等价，其中两个直和有相同的项数.

78

(c) $\begin{bmatrix} A & & & 0 \\ & B & & \\ & & \ddots & \\ 0 & & & B \end{bmatrix}$和$\begin{bmatrix} A & & & 0 \\ & C & & \\ & & \ddots & \\ 0 & & & C \end{bmatrix}$酉等价，其中两个直和有相同的项数.

7. 证明，诸某个次数 k 有 2^{k-1}个不同的形如(2.2.5)的字 $\mathbf{W}(s, t)$. 因而得出至多有 $4n^2$ 个次数至多为 $2n^2$ 的不同的字.

8. 试给出两个 2×2 矩阵，这两个矩阵满足恒等式(2.2.2)，但它们不酉等价. 解释其原因.

进一步阅读与注释 关于定理(2.2.6)的原始证明可参看 W. Specht, "Zur Theorie der Matrizen II," *Jahresbericht der Deutschen Mathematiker Vereinigung* 50(1940), 19-23. 定理(2.2.8)的证明可见 C. Pearcy, "A Complete Set of Unitary Invariants for Operators Generating Finite W^*-Algebras of Type I," *Pacific J. Math.* 12(1962), 1405-1416. 关于低阶矩阵酉等价的讨论可见 C. Pearcy, "A Complete Set of Unitary Invariants for 3×3 Complex Matrices," *Trans. Amer. Math. Soc.* 104(1962), 425-429.

2.3 Schur 酉三角化定理

初等矩阵理论最有用的基本事实或许是任一矩阵 $A\in M_n$ 酉等价于一个上三角矩阵 T[也酉等价于一下三角矩阵]. T 的诸对角元自然是 A 的特征值. 尽管这种形式不唯一，但是它是在酉等价下所能得到的最简形式.

2.3.1 定理(Schur) 已知 $A\in M_n$ 有特征值 $\lambda_1,\cdots,\lambda_n$，它们按任意规定的次序排列，那么存在一个酉矩阵 $U\in M_n$，使得

$$U^*AU=T=[t_{ij}]$$

是具有对角元 $t_{ii}=\lambda_i$，$i=1,\cdots,n$ 的上三角矩阵，即每个方阵 A 酉等价于其对角元依次是 A 的特征值的三角矩阵. 此外，如果 $A\in M_n(\mathbf{R})$，且 A 的所有特征值都是实数，那么，可选择 U 为实正交矩阵.

证明： 证明本身就是实施一系列相同形式的化简算法并得到化简结果. 设 $x^{(1)}$ 是 A 的相应于征特值 λ_1 的正规化特征向量. 非零向量 $x^{(1)}$ 可以扩充为 $\mathbf{C}^n$ 的一个基

$$x^{(1)},y^{(2)},y^{(3)},\cdots,y^{(n)}.$$

79

应用 Gram-Schmidt 标准正交化过程(0.6.4)于这个基，便得到 $\mathbf{C}^n$ 的一个标准正交基

$$x^{(1)},z^{(2)},\cdots,z^{(n)}.$$

从左至右把这些标准正交向量排成酉矩阵 U_1 的诸列. 因为 AU_1 的第1列是 $\lambda_1x^{(1)}$，计算表明，$U_1^*(AU_1)$有形式

$$U_1^*AU_1=\begin{bmatrix}\lambda_1 & *\\ 0 & A_1\end{bmatrix}$$

矩阵 $A_1\in M_{n-1}$ 有特征值 $\lambda_2,\cdots,\lambda_n$. 设 $x^{(2)}\in\mathbf{C}^{n-1}$ 是 A_1 的相应于 λ_2 的正规化特征向量，然后完全重复上述步骤. 确定一个酉矩阵 $U_2\in M_{n-1}$，使得

$$U_2^*A_1U_2=\begin{bmatrix}\lambda_2 & *\\ 0 & A_2\end{bmatrix},$$

并设

$$V_2=\begin{bmatrix}1 & 0\\ 0 & U_2\end{bmatrix}.$$

于是，矩阵 V_2 和 U_1V_2 是酉矩阵，因而 $V_2^*U_1^*AU_1V_2$ 有形式

$$V_2^*U_1^*AU_1V_2=\begin{bmatrix}\lambda_1 & * & \\ 0 & \lambda_2 & *\\ 0 & & A_2\end{bmatrix}.$$

继续作这种化简便到酉矩阵 $U_i\in M_{n-i+1}$，$i=1$，…，$n-1$ 和酉矩阵 $V_i\in M_n$，$i=2$，…，$n-1$. 矩阵

$$U=U_1V_2V_3\cdots V_{n-1}$$

是酉矩阵，而 U^*AU 给出了所要求的形式.

如果 $A\in M_n(\mathbf{R})$ 的所有特征值恰好是实数，那么，相应的特征向量可以选实向量，且上述所有步骤可以用实的算术运算来完成，这就是证明了后一个论断. □

附注　仿照(2.3.1)的证明便可看出，在定理的叙述中，可用“下三角”代替“上三角”，当然它对应于一个不同的酉等价 U.

[80] **2.3.2 例**　不论是酉矩阵 U，还是定理(2.3.1)中的三角矩阵 T，都不是唯一的. 不仅 T 的对角元(A 的特征值)可以依任何顺序出现，而且酉等价的上三角矩阵在其对角线上方可以呈现完全不同的形式. 例如，

$$T=\begin{bmatrix}1&1&4\\0&2&2\\0&0&3\end{bmatrix}\text{和 }T_2=\begin{bmatrix}2&-1&3\sqrt{2}\\0&1&\sqrt{2}\\0&0&3\end{bmatrix}$$

是经

$$U=\frac{1}{\sqrt{2}}\begin{bmatrix}1&1&0\\1&-1&0\\0&0&\sqrt{2}\end{bmatrix}$$

的酉等价. 一般说来，许多不同的上三角矩阵可以在同一个酉等价类之中.

附注　应指出的是，证明(2.3.1)的技巧不过是如 1.4 节中习题 8 所概述的顺序压缩技巧.

练习　如果 $A=[a_{ij}]$ 和 $B=[b_{ij}]\in M_2$ 相似，且 $\sum_{i,j}|a_{ij}|^2=\sum_{i,j}|b_{ij}|^2$，证明 A 和 B 酉等价. 用例子说明这在高维情形不成立. **提示**：注意到，如果 A 和 B 酉等价，那么 $A+A^*$ 和 $B+B^*$ 也酉等价. 考虑

$$A=\begin{bmatrix}1&3&0\\0&2&4\\0&0&3\end{bmatrix}\text{和 }B=\begin{bmatrix}1&0&0\\0&2&5\\0&0&3\end{bmatrix}.$$

(2.3.1)的有用推论是，矩阵交换族可以同时上三角化.

2.3.3 定理　设 $\mathscr{F}\subseteq M_n$ 是交换族. 那么存在一个酉矩阵 $U\in M_n$ 使得对每个 $A\in\mathscr{F}$，U^*AU 是上三角的.

证明：同到(2.3.1)的证明. 在原证明的每一步运用(1.3.17)，在每一步都可选定一个特征向量(和酉矩阵)，且对每个 $A\in\mathscr{F}$ 都可以选定这同一个特征向量(和酉矩阵). 又酉等价保持

[81] 交换性，且分块矩阵乘法计算表明，如果形如

$$\begin{bmatrix}A_{11}&A_{12}\\0&A_{22}\end{bmatrix}\text{和}\begin{bmatrix}B_{11}&B_{12}\\0&B_{22}\end{bmatrix}$$

两个矩阵可交换，那么 A_{22} 和 B_{22} 也可交换．因此，在证明(2.3.1)的化简过程中的每一步，每个 A_i 都继承了交换族性质．我们得出，对交换族的所有成员，可用相同的方式选择(2.3.1)的 U 中的所有组成部分，这就证明了(2.3.3)．值得指出的是，这里并没有断言，对各个族的成员的特征值，可以选取任一特定的次序．它取定的只是应用(1.3.17)时所得到的那个次序． □

下面的定理包括了(2.3.1)的严格的实形式．

2.3.4 定理 如果 $A\in M_n(\mathbf{R})$，那么有实正交矩阵 $Q\in M_n(\mathbf{R})$，使得

$$Q^TAQ=\begin{bmatrix}A_1 & & & * \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_k\end{bmatrix}\in M_n(\mathbf{R}),\quad 1\leqslant k\leqslant n, \tag{2.3.5}$$

其中每个 A_i 或是 1×1 实矩阵或是具有一对非实的复共轭特征值的 2×2 实矩阵．对角子块 A_i 可以按任意的次序排列．

一般说来，不能指望通过一个实相似(更不用说实正交相似了)把一个实矩阵化成上三角形式，因为可能的话，对角元将是特征值，而特征值可以不是实数．形式(2.3.5)是通过实正交相似所能得到的与三角形式最接近的形式．如果 A 有任何非实特征值，形式(2.3.5)就不会是上三角矩阵，但它总是呈上 Hessenberg 形状．

练习 试修改(2.3.1)的论证以证明(2.3.4)．**提示**：如果 λ 是实矩阵 A 的实特征值，那么有一个相应的实特征向量，可以用它来压缩 A，使呈(2.3.1)中的形状．如果 $\lambda=\alpha+i\beta$ 是 A 的非实特征值，且 $Ax=\lambda x$，$x=u+iv\neq0$，u，$v\in\mathbf{R}^n$，证明 $Au=\alpha u-\beta v$，$Av=\alpha v+\beta u$ 和 $A\bar{x}=\bar{\lambda}\bar{x}$，再证明 $\{x,\bar{x}\}$ 是无关组．证明 $\{u,v\}$ 是无关组，然后对它应用 Gram-Schmidt 过程得到实的标准正交组 $\{w,z\}$．设 Q_1 是前两列为 w 和 z 的实正交矩阵，证明

[82]

$$Q_1^TAQ_1=\left[\begin{array}{cc|c} * & * & \\ * & * & * \\ \hline \multicolumn{2}{c|}{0} & \tilde{A}\end{array}\right].$$

这时，一次可使 A 压缩两列．应该指出，在(2.3.5)中，相应于每个实特征值和每对复共轭特征值的子块 A_i 可以按任意规定的顺序排列．

(2.3.3)也有一个在实的情形下的变形．

2.3.6 定理 设 $\mathcal{F}\subseteq M_n(\mathbf{R})$ 是交换族．则存在实正交矩阵 $Q\in M_n(\mathbf{R})$，使得对每个 $A\in\mathcal{F}$，Q^TAQ 为(2.3.5)的形式．

练习 试修改(2.3.3)的论证以证明(2.3.6)．**提示**：首先，用所有实的公共特征向量压缩 $\mathcal{F}$ 的所有成员．然后考察非实公共特征向量，且如(2.3.4)的证明所做的那样一次压缩两列．要注意的是，经一个公共的实正交相似之后，$\mathcal{F}$ 的不同成员可以有不同的 2×2 对角子块，但是，如果一个成员在某个位置有一个 2×2 子块，而另一个成员却没有，那么后者必须在相应位置有一对相同的 1×1 子块．

习题

1. 设 $x\in\mathbf{C}^n$ 是给定的单位向量($x^*x=1$)，且记 $x=[x_1,\ y^T]^T$，其中 $x_1\in\mathbf{C}$，而 $y\in\mathbf{C}^{n-1}$．

选取 $\theta\in\mathbf{R}$ 使 $e^{i\theta}x_1\geqslant 0$．然后定义 $z=e^{i\theta}x=[z_1,\ \zeta^T]$，其中 $z_1\in\mathbf{R}$ 是非负实数，而 $\zeta\in\mathbf{C}^{n-1}$．证明矩阵

$$V=\left[\begin{array}{c|c} z_1 & \zeta^* \\ \hline \zeta & -I+\frac{1}{1+z_1}\zeta\zeta^* \end{array}\right]$$

是酉矩阵．提示：计算 $V^*V=V^2$．得出矩阵 $U=e^{-i\theta}V=[xu_2\cdots u_n]$是其第一列为已知向量 x 的酉矩阵．这为 Schur 定理(2.3.1)的证明中的逐个压缩步骤，提供了一个求所需酉矩阵的构造性方法．

83

2. 如果 $x\in\mathbf{R}^n$ 是给定的单位向量，说明如何改进习题 1 中所描述的构造法，以便得到第一列是 x 的实正交矩阵 $Q\in M_n(\mathbf{R})$．证明你的构造方法是可行的．

3. 设 $A\in M_n(\mathbf{R})$．解释 A 的非实特征值(如果有的话)必须成共轭对出现．

4. 考虑族

$$\mathscr{F}=\left\{\begin{bmatrix}0 & -1\\0 & -1\end{bmatrix},\begin{bmatrix}1 & 1\\0 & -1\end{bmatrix}\right\},$$

试说明定理(2.3.3)中的交换性假设虽然是 $\mathscr{F}$ 可同时酉上三角化的充分条件，但它不是必要条件．

5. 设 $\mathscr{F}=\{A_1,\ \cdots,\ A_k\}\subset M_n$ 是给定的族，且设

$$\mathscr{G}=\{A_iA_j:i,j=1,2,\cdots,k\}$$

是 $\mathscr{F}$ 中矩阵的所有两两乘积组成的族．事实上，如果 $\mathscr{G}$ 是交换族，那么，$\mathscr{F}$ 可以同时酉上三角化的充分必要条件是每个换位子 $A_iA_j-A_jA_i$ 的每个特征值都是零．证明关于 $\mathscr{G}$ 的交换性假设是比 $\mathscr{F}$ 的交换性假设要弱的假设．同时证明，习题 4 中的族 $\mathscr{F}$ 有一个相应的交换族 $\mathscr{G}$，它还满足零特征值条件．

6. 设 $A,\ B\in M_n$ 已知，且假设 A 和 B 同时相似于上三角矩阵；即对某个非奇异矩阵 $S\in M_n$，$S^{-1}AS$ 和 $S^{-1}BS$ 都是上三角矩阵．证明 $AB-BA$ 的每个特征值必定是零．提示：如果 $\Delta_1,\ \Delta_2\in M_n$ 都是上三角矩阵 $\Delta_1\Delta_2-\Delta_2\Delta_1$ 的主对角线是什么？

7. 虽然每个方阵可经酉相似化成上三角形式，但这对复正交相似不成立．如果某个 $A\in M_n$ 可以写成 $A=Q\Delta Q^T$，其中，$Q\in M_n$ 是复正交矩阵，而 $\Delta\in M_n$ 是上三角矩阵，证明 A 至少有一个特征向量 $x\in\mathbf{C}^n$ 使 $x^Tx\neq 0$．考察 $A=\begin{bmatrix}1 & i\\ i & -1\end{bmatrix}$，试说明不是每个 $A\in M_n$ 可以经一个复正交相似上三角化．

8. 设 $Q\in M_n$ 是给定的复正交矩阵，且假定 $x\in\mathbf{C}^n$ 是 Q 的相应于特征值$\lambda\neq\pm 1$ 的特征向量．证明 $x^Tx=0$．提示：在恒等式 $Qx=\lambda x$ 的两边各乘以自己的转置．关于两个特征值都不同于± 1 的 2×2 复正交矩阵族的例子，可参看(2.1)节的习题 8(a)．证明，这些矩阵中没有任何矩阵可经正交相似化为上三角矩阵．

84

进一步阅读 关于习题 5 所确立的定理(2.3.3)的较强形式，其证明可参看 Y. P. Hong and R. A. Horn, "On Simultaneous Reduction of Families of Matrices to Triangular or Diagonal Form by Unitary Congruences," *Linear and Multilinear Algebra* 17(1985), 271-288.

2.4 Schur 定理的若干推论

Schur 酉三角化的几个基本推论说明了它的效用.

练习 利用(2.3.1)证明，如果 $A\in M_n$ 有特征值 $\lambda_1, \cdots, \lambda_n$，(包括重特征值)，那么 $\det A=\prod_{i=1}^{n}\lambda_i$ 和 $\operatorname{tr} A=\sum_{i=1}^{n}\lambda_i$. 在第 1 章中，这是用另外的方法证明的. **提示**：关于迹，应想到由直接计算可推出 $\operatorname{tr} AB=\operatorname{tr} BA$，因而迹是相似不变量. 关于特征值的其他初等对称函数，你认为如何?

每个矩阵都满足它自己的特征方程(2.4.2)，这个结论可由 Schur 定理和关于三角矩阵乘法的一个简单结果推出.

2.4.1 引理 假定 $R=[r_{ij}]$ 和 $T=[t_{ij}]\in M_n$ 是上三角矩阵，且 $r_{ij}=0$，$1\leqslant i, j\leqslant k<n$，$t_{k+1,k+1}=0$. 设 $T'=[t'_{ij}]=RT$，那么 $t'_{ij}=0$，$1\leqslant i, j\leqslant k+1$.

证明：因为 $R(1, 2, \cdots, k)=0$ 和 $t_{k+1,k+1}=0$，所以 R 和 T 有形状

$$R=\left[\begin{array}{c|c} 0 & * \\ \hline 0 & \begin{matrix} * & & * \\ & \ddots & \\ 0 & & * \end{matrix} \end{array}\right],\quad T=\left[\begin{array}{c|c} \begin{matrix} * & & * \\ & \ddots & \\ 0 & & * \end{matrix} & * \\ \hline 0 & \begin{matrix} 0 & & & \\ & * & & * \\ & & \ddots & \\ 0 & 0 & & * \end{matrix} \end{array}\right],$$

其中，分块后的右上子块都是 $k\times k$ 的. 根据分块矩阵的乘法[见(0.7)的有关记号和基本结果]，T'的 $k\times k$ 左上子块显然是 0. 通过观察发现，R 的前 $k+1$ 行有 0 的位置对着 T 的第 $k+1$ 列的所有非零位置，而 T 的前 $k+1$ 列有 0 的位置对着 R 的第 $k+1$ 行的所有非零位置. 于是矩阵乘法表明，T'(按同样的分块)有形状 85

$$T'=\left[\begin{array}{c|c} 0 & \begin{matrix} 0 & \\ \vdots & * \\ 0 & \end{matrix} \\ \hline 0 & \begin{matrix} 0 & & \\ & * & * \\ \vdots & & \ddots \\ 0 & 0 & * \end{matrix} \end{array}\right],$$

因而 $T'(\{1, \cdots, k+1\})=0$. 这正是我们要证明的. □

练习 证明两个上三角矩阵的乘积是上三角矩阵，并证明两个有相同划分的分块上三角矩阵的乘积是分块上三角矩阵.

练习 推广(2.4.1)，证明，如果 R 和 T 是上三角矩阵，且 $T'=RT$，那么，

$$T'(\{i,i+1,\cdots,i+j\}) = R(\{i,i+1,\cdots,i+j\})T(\{i,i+1,\cdots,i+j\}).$$

2.4.2 定理(Cayley-Hamiltom) 设 $p_A(t)$是 $A\in M_n$ 的特征多项式. 那么

$$p_A(A) = 0.$$

证明：因为 $p_A(t)$是 n 次的，且首系数为 1，又 $p_A(t)=0$ 的根正是 A 的特征值λ_1，…，λ_n(计相重特征值)，可以把 $p_A(t)$分解成

$$p_A(t) = (t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n).$$

利用(2.3.1)，把 A 表成

$$A = UTU^*,$$

其中 T 是上三角矩阵，且在它的第 i 个对角元位置有 λ_i，$i=1$，…，n. 现在算出

$$\begin{aligned} p_A(A) &= p_A(UTU^*) = (UTU^*-\lambda_1 I)(UTU^*-\lambda_2 I)\cdots(UTU^*-\lambda_n I) \\ &= [U(T-\lambda_1 I)U^*][U(T-\lambda_2 I)U^*]\cdots[U(T-\lambda_n I)U^*] \\ &= U[(T-\lambda_1 I)(T-\lambda_2 I)\cdots(T-\lambda_n I)]U^* \\ &= Up_A(T)U^*, \end{aligned}$$

86

并注意到，$p_A(A)=0$，当且仅当 $p_A(T)=0$. 此外，由引理(2.4.1)得出 $p_A(T)=0$. $T-\lambda_1 I$ 的左上 1×1 子块是 0，而 $T-\lambda_2 I$ 的 2，2 元是 0；因为它们都是上三角矩阵，所以$(T-\lambda_1 I)(T-\lambda_2 I)$的左上 2×2 子块是 0，根据归纳推理，因为$(T-\lambda_1 I)\cdots(T-\lambda_k I)$的左上 $k\times k$ 子块和$(T-\lambda_{k+1} I)$的 $k+1$，$k+1$ 元各为 0. 且都是上三角矩阵，所以$(T-\lambda_1 I)\cdots(T-\lambda_{k+1} I)$的左上$(k+1)\times(k+1)$子块是 0. 一直连续计算到 n，便得到乘积 $p_A(T)=(T-\lambda_1 I)\cdots(T-\lambda_n I)=0$. 证毕. □

练习 下述关于命题 $p_A(A)=0$ 的论证有什么错误? "因为对于 $A\in M_n$ 的每个特征值 λ 有 $p_A(\lambda)=0$，又因为 $q(A)$的特征值是 $q(\lambda)$，其中 q 是一个多项式，由此推出 $p_A(A)$的所有特征值是 0. 因此，$p_A(A)$是 0."这是关于 Cayley-Hamilton 定理的一个常见的错误证明. 给出了一个明显的例子来说明它错在哪里.

练习 下述论证的错误是什么? "因为 $p_A(t)=\det(tI-A)$，$p_A(A)=\det(AI-A)=\det(A-A)=\det 0=0$，所以 $p_A(A)=0$".

如果 $p_A(t)=\det(tI-A)$表示 $A\in M_n$ 的特征多项式，那么特征方程是 $p_A(t)=0$. 特征方程的根是 A 的特征值. Cayley-Hamilton 定理常常解释为"每个方阵都满足自己的特征方程，"但这必须认真弄懂：纯量多项式首先是作为 $p_A(t)=\det(tI-A)$计算出来的，然后才从特征多项式出发作出矩阵 $p_A(A)$.

我们已经对具有复分量的矩阵证明了 Cayley-Hamilton 定理. 因而它对分量取自复数域的任一子域(例如，实数域或有理数域)的矩阵也必定成立. 实际上，Cayley-Hamilton 定理纯粹是一个形式结果，它对分量取自任何域的或更一般地，取自任何交换环的矩阵也成立.

Cayley-Hamilon 定理的一个重要应用是，可以把 $A\in M_n$ 的幂 A^k $(k\geqslant n)$写成 I，A，A^2，…，A^{n-1}的线性组合. 根据线性相关的理论，容易证明，幂 A^{n^2} 和 A 的更高次幂可以表成较低次幂的线性组合(因为若把 M_n 看作复数域上的向量空间，M_n 的维数是 n^2)，但是 Cayley

Hamilton 定理是上述结果的显著改进. [87]

2.4.3 例 设

$$A=\begin{bmatrix}3 & 1\\ -2 & 0\end{bmatrix},$$

则 $p_A(t)=t^2-3t+2$，并且 $A^2-3A+2I=0$. 因此，$A^2=3A-2I$；$A^3=A(A^2)=3A^2-2A=3(3A-2I)-2A=7A-6I$；$A^4=7A^2-6A=15A-14I$，等等. 另外，因为 $p_A(t)$中的常数项，即 A 的行列式非零，所以 A 非奇异，且可以把 A^{-1}写成 A 的一个多项式. 再由 $p_A(A)=A^2-3A+2I=0$，得到 $2I=-A^2+3A=A(-A+3I)$，或

$$I=A\left[\frac{1}{2}(-A+3I)\right].$$

这表明 $A^{-1}=-\frac{1}{2}A+\frac{3}{2}I=\begin{bmatrix}0 & -1/2\\ 1 & 3/2\end{bmatrix}$.

练习 已知 $A\in M_n$ 有特征多项式

$$p_A(t)=t^n+a_{n-1}t^{n-1}+a_{n-2}t^{n-2}+\cdots+a_1t+a_0$$

试把 A^n 写成关于 A 的次数至多是 $n-1$ 的多项式. 试对 A 的次数大于 $n-1$ 的几个相邻次数的幂，把它们也写成 A 的次数至多是 $n-1$ 的多项式. 另外，假定 A 非奇异($a_0\neq 0$)，把 A^{-1}也写成 A 的次数至多是 $n-1$ 的多项式. 这后一结论可作为(2.4.2)的一个推论.

2.4.4 推论 如果 $A\in M_n$ 是非奇异矩阵，那么存在次数至多是 $n-1$ 的多项式 $q(t)$(它的系数取决于 A)，使得 $A^{-1}=q(A)$.

练习 如果两个矩阵 A，$B\in M_n$ 相似，证明其中一个矩阵的任一多项式相似于另一矩阵的同一多项式，特别是，一个矩阵所满足的任一多项式方程，另一个矩阵也满足. 试考虑逆命题：满足同一组多项式的两个同阶矩阵是相似的. 它成立还是不成立?

2.4.5 例 已经证明，每个矩阵 $A\in M_n$ 满足某个 n 次多项式方程. 例如，特征方程就是一个例子. 但是，$A\in M_n$ 满足一个次数小于 n 的多项式方程也是可能的. 设

$$A=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\in M_3,$$

则 A 满足 $q(A)=0$，其中 $q(t)=t^2-2t+1$ 是 2 次的. [88]

练习 证明一个可对角化矩阵满足一个次数等于其不同特征值的个数的多项式方程，并且它不满足更低次的多项式方程. 一个矩阵所满足的(首项系数为 1 的)极小次数的多项式(它的极小多项式)是下一章要进一步研究的对象，它与 Jordan 标准形有密切的关系. **提示**：考虑 $q(t)=(t-\lambda_1)\cdots(t-\lambda_k)$，其中 $\lambda_i\neq\lambda_j$.

Schur 定理的另一个用途是，它们能用两种说法来解释每个矩阵“几乎”是可对角化的. 第一种说法是，存在一个可对角化矩阵，它可以任意接近于一个已知矩阵；第二种说法是，任一已知矩阵相似于这样一个上三角矩阵，它的非对角元可以任意地小.

2.4.6 定理 设 $A=[a_{ij}]\in M_n$. 对任意 $\epsilon>0$，存在一个矩阵 $A(\epsilon)=[a_{ij}(\epsilon)]\in M_n$，它有 n 个不同的特征值(因而它是可对角化的)，且使得

$$\sum_{i,j=1}^{n}|a_{ij}-a_{ij}(\epsilon)|^2<\epsilon.$$

证明：设 $U\in M_n$ 是酉矩阵，使得 $U^*AU=T$ 是上三角矩阵. 设 $E=\mathrm{diag}(e_1, e_2, \cdots, e_n)$，其中 $e_1, \cdots, e_n$ 选择这样一些数，使得

$$|e_i|<\left(\frac{\epsilon}{n}\right)^{1/2}$$

且数 $t_{11}+e_1, t_{22}+e_2, \cdots, t_{nn}+e_n$ 是互不相同的. (稍加思考便会发现，这是可以做到的.)于是 $T+E$ 有 n 个不同的特征值：$t_{11}+e_1, \cdots, t_{nn}+e_n$，且 $A+UEU^*$ 也如此，这是因为它相似于 $T+E$. 设 $A(\epsilon)=A+UEU^*$，从而 $A-A(\epsilon)=-UEU^*$，且

$$\sum_{i,j}|a_{ij}-a_{ij}(\epsilon)|^2=\sum_{i=1}^{n}|e_i|^2<n\left(\frac{\epsilon}{n}\right)=\epsilon.$$

我们已经用到了(2.2.2). 因此，$A(\epsilon)$适合定理的要求. □

练习 证明(2.4.6)中条件 $\sum_{i,j}|a_{ij}-a_{ij}(\epsilon)|^2<\epsilon$ 可以用 $\max_{i,j}|a_{ij}-a_{ij}(\epsilon)|<\epsilon$ 代替. **提示**：运用上述定理时，用 ϵ^2 代替 ϵ，并意识到，如果平方和小于 ϵ^2，那么其中每一项的绝对值必小于 ϵ.

[89] **2.4.7 定理** 设 $A\in M_n$. 对任意 $\epsilon>0$，存在一个非奇异矩阵 $S_\epsilon\in M_n$，使得

$$S_\epsilon^{-1}AS_\epsilon=T_\epsilon=[t_{ij}(\epsilon)]$$

是上三角矩阵，且 $|t_{ij}(\epsilon)|<\epsilon$，$1\leqslant i<j\leqslant n$.

证明：首先应用 Schur 定理得到酉矩阵 $U\in M_n$ 和上三角矩阵 $T\in M_n$，使得

$$U^*AU=T$$

对一个非零纯量 α，定义 $D_\alpha=\mathrm{diag}(1, \alpha, \alpha^2, \cdots, \alpha^{n-1})$且令 $t=\max_{i<j}|t_{ij}|$. 假定 $\epsilon<1$，因为在这种情形下证明本定理就可以了. 如果 $t\leqslant 1$，设 $S_\epsilon=UD_\epsilon$；如果 $t>1$；设 $S_\epsilon=UD_{1/t}D_\epsilon$. 不论在哪种情形，相应的 S_ϵ 都可证明定理的论断. 例如，如果 $t\leqslant 1$，通过简单的计算可知，$t_{ij}(\epsilon)=t_{ij}\epsilon^{-i}\epsilon^{j}=t_{ij}\epsilon^{j-i}$，它的绝对值不大于 ϵ^{j-i}，又如果 $i<j$，这个值就不大于 ϵ. 另一方面，如果 $t>1$，仅仅用一个经 $D_{1/t}$的相似作用于矩阵 T，便得到其所有非零元的绝对值不超过 1 的矩阵. □

练习 证明(2.4.7)的下述变形：若 $A\in M_n$ 且 $\epsilon>0$，则存在一个非奇异的 $S_\epsilon\in M_n$，使得 $S_\epsilon^{-1}AS_\epsilon=T_\epsilon=[t_{ij}(\epsilon)]$是上三角矩阵，且$\sum_{i<j}|t_{ij}(\epsilon)|<\epsilon$. **提示**：应用(2.4.7)，其中用$[2/n(n-1)]\epsilon$代替$\epsilon$.

从 Schur 定理出发不难证明它的一个推广，这个推广向下一章中要出现的 Jordan 标准形迈出了重要的一步.

2.4.8 定理 假定 $A\in M_n$ 有 n_i 重特征值λ_i，$i=1, \cdots, k$，且 $\lambda_1, \cdots, \lambda_k$ 是互不相同的. 那么 A 相似于形如

$$\begin{bmatrix} T_1 & & & 0 \\ & T_2 & & \\ & & \ddots & \\ 0 & & & T_k \end{bmatrix}$$

的矩阵，其中 $T_i \in M_{n_i}$ 是所有对角元为 λ_i 的上三角矩阵，$i=1, \cdots, k$. 如果 $A \in M_n(\mathbf{R})$，且 A 的所有特征值都是实数，那么也有同样的结果，且相似矩阵可以取实矩阵.

证明：首先应用 Schur 定理(2.3.1)使 A(酉)相似于一个上三角矩阵 $T=[t_{ij}]$，并且假定在 T 的对角线上作了编排，使所有的项 λ_1 都出现在前面，其次是项 λ_2，如此等等. 下面，要对 T 作一系列简单的(非酉)相似变换，以期得到对角线上方的各个 0 元，且不改变 T 的对角或者上三角结构. 设 E_{rs} 是 M_n 中的在 r，s 位置是 1，而其他位置都为 0 的矩阵. 注意，对 $r \neq s$ 和任一纯量 α，$I+\alpha E_{rs}$ 是非奇异矩阵，且 $(I+\alpha E_{rs})^{-1}=I-\alpha E_{rs}$. 此外，由直接计算可知，对 $r<s$，经 $I+\alpha E_{rs}$ 的相似

[90]

$$(I+\alpha E_{rs})^{-1}T(I+\alpha E_{rs}) = (I-\alpha E_{rs})T(I+\alpha E_{rs})$$

只改变 T 的第 r 行中位于第 s 列右边的元素和 T 的第 s 列中位于第 r 行上方的元素，

$$\left[\begin{array}{c} \uparrow \\ \llcorner\!\!\rightarrow \\ r,s \end{array}\right],$$

并且用

$$t_{rs}+\alpha(t_{rr}-t_{ss})$$

代替 t_{rs}. 因此，如果 $t_{rr} \neq t_{ss}$，选取

$$\alpha = \frac{-t_{rs}}{(t_{rr}-t_{ss})}$$

便可使 r，s 元为 0，而对有关结构未作其他的变更. 现在考虑 T 中一系列位置：$(n-1, n)$；$(n-2, n-1)$，$(n-2, n)$；$(n-3, n-2)$，$(n-3, n-1)$，$(n-3, n)$；$(n-4, n-3)$，…，如果 $t_{rr} \neq t_{ss}$，可以依次指定一种相似使这些位置的每个元素变为 0. 我们注意到，这样做不会影响已经变成 0 的元. 所得到的矩阵将相似于 A. 这正是所要求的形式. □

练习 证明，如果 $A \in M_n(\mathbf{R})$，且它的所有特征值都是实数，那么，在(2.4.8)的证明中所必需的运算都可以在实数范围内完成，于是，在这种情形，定理保证能得到分块对角矩阵，而所必需的相似矩阵可以取实矩阵.

附注 假定一个已知矩阵 $A \in M_n$ 是上三角矩阵，且可以假定它已化简成如下形式(如果有必要，可对它作置换相似)

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1k} \\ & \ddots & \vdots \\ 0 & & A_{kk} \end{bmatrix},$$

[91]

其中，每个对角子块 A_{ii} 是上三角矩阵，且在其对角线上只有 λ_i，还假定 $i \neq j$ 时有 $\lambda_i \neq \lambda_j$，证明定理(2.4.8)时所使用的算法说明，A 相似于

$$\begin{bmatrix} A_{11} & & 0 \\ & \ddots & \\ 0 & & A_{kk} \end{bmatrix}.$$

即：在这种情形下，所有非对角子块可以用零子块来代替，且保持相似性. 因为酉相似保持各元素绝对值的平方和，所以应当注意，如果有任一非对角子块 A_{ij} 是非零的，就不可能用一个酉相似得到这一结果.

现在利用 Schur 定理的交换族形式(2.3.3)证明，对可交换的矩阵，特征值可(按某个顺序)"相加".

2.4.9 定理 设 A，$B\in M_n$ 分别有特征值 α_1，…，α_n 和 β_1，…，β_n. 如果 A 和 B 可交换，那么存在指标 1，…，n 的一个排列 i_1，…，i_n，使得 $A+B$ 的特征值是 $\alpha_1+\beta_{i_1}$，$\alpha_2+\beta_{i_2}$，…，$\alpha_n+\beta_{i_n}$. 特别是，如果 A 与 B 可交换，那么 $\sigma(A+B)\subseteq\sigma(A)+\sigma(B)$.

证明：如果 A 与 B 可交换，根据(2.3.3)，它们可以同时上三角化，即存在酉矩阵 $U\in M_n$，使得

$$U^*AU=T \text{ 和 } U^*BU=R$$

都是上三角矩阵，且分别具有对角元 α_1，…，α_n 及 β_{i_1}…，β_{i_n}.

$$U^*(A+B)U=T+R$$

有对角元

$$\alpha_1+\beta_{i_1},\alpha_2+\beta_{i_2},\cdots,\quad \alpha_n+\beta_{i_n}.$$

因而也以它们为特征值. 因为 $A+B$ 相似于 $T+R$，所以它们必定也是 $A+B$ 的特征值. □

2.4.10 例 应注意的是，即使 A 与 B 可交换，也未必所有形如 $\alpha_i+\beta_j$ 的成员都是 $A+B$ 的特征值. 考察对角矩阵

92

$$A=\begin{bmatrix}1 & 0\\0 & 2\end{bmatrix} \quad 和 \quad B=\begin{bmatrix}3 & 0\\0 & 4\end{bmatrix}$$

便会发现，$1+4=5\notin\{4,6\}=\sigma(A+B)$. 因此，当 A 与 B 可交换时，$\sigma(A+B)$ 包含在 $\sigma(A)+\sigma(B)$ 中，但一般不相等.

2.4.11 例 如果 A 与 B 不交换，那么要用 $\sigma(A)+\sigma(B)$ 来说明 $\sigma(A+B)$ 是很困难的，特别是，$\sigma(A+B)$ 不一定包含在 $\sigma(A)+\sigma(B)$ 中. 设

$$A=\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} \quad 和 \quad B=\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix},$$

则 $\sigma(A+B)=\{-1,1\}$，而 $\sigma(A)=\sigma(B)=\{0\}$.

2.4.12 例 (2.4.9)的逆命题成立吗？如果 A 和 B 的特征值可按某个顺序相加，A 与 B 一定可交换吗？回答是否定的，即便是对于所有纯量 α 和 β，αA 和 βB 的特征值可按某个顺序相加，A 与 B 也未必可交换. 这是一种有趣的现象，而刻划这样一对矩阵的特征还是一个尚未解决的问题！设

$$A=\begin{bmatrix}1&4&5\\0&2&6\\0&0&3\end{bmatrix}\quad 和\quad B=\begin{bmatrix}2&1&2\\0&3&3\\0&0&4\end{bmatrix},$$

它们的特征值可相加，但 A 与 B 不交换．显然，可同时上三角化是特征值具有可加性的充分条件，但它又不是必要的．自然，上三角矩阵不一定可交换．

2.4.13 推论 假定 A，$B\in M_n$ 是分别具有特征值 α_1，…，α_n 和 β_1，…，β_n 的可交换矩阵．如果 $\alpha_i\neq-\beta_j$，i，$j=1$，…，n，那么 $A+B$ 是非奇异矩阵．

练习 用(2.4.9)证明(2.4.13)．

练习 证明，对任意一对 A，$B\in M_n$(可交换或不可交换)，$A+B$ 的各特征值的和是 A 的各特征值的和加上 B 的各特征值的和．**提示：** $\operatorname{tr}(A+B)$是什么?

我们已考察过可对角化矩阵的同时对角化问题，对此，交换性是一个容易验证的必要充分条件，也考察过同时上三角化问题，对于它，交换性是一个充分条件，但不是必要条件．鉴于能够证明两个已知矩阵不可同时三角化往往是很有用的，所以，希望能够找到一个比特征值具有可加性这一条件更强的必要条件．下面的例子指出了实现这一条件的途径． 93

2.4.14 例 设

$$A=\begin{bmatrix}0&1&0\\0&0&1\\0&0&0\end{bmatrix}\quad 和\quad B=\begin{bmatrix}0&0&0\\1&0&0\\0&1&0\end{bmatrix}.$$

A 和 B 都有三重特征值 0．它们的任意线性组组合 $aA+bB$ 也是如此，于是，它们的特征值可相加，由于这些原因，似乎有理由相信 A 和 B 是可同时三角化的．但是，假如有某个非奇异矩阵 $S\in M_3$，使得 SAS^{-1}和 SBS^{-1}都是上三角矩阵，那么$(SAS^{-1})(SBS^{-1})=SABS^{-1}$的特征值一定是 A 和 B 的特征值按某个顺序的乘积．可是，AB 的特征值的集合是$\{-1, 0, 1\}$，它并不包含在集$\{0\}$和集$\{0\}$的集乘积之中．由此，我们得出结论，A 和 B 是不可同时上三角化的．

练习 证明上例中的一个论断：如果 C，$D\in M_n$ 都是上三角矩阵，那么 CD 的特征值是 C 和 D 的特征值按某个顺序的乘积；即 $\sigma(CD)\subseteq\sigma(C)\sigma(D)$．

给一个不一定是酉相似的同时上三角化问题[不过要参看(2.6)节]可以用下面的 McCoy 定理来完整地描述，我们略去了它的证明．请回忆一下曾述及任意多个变元的多项式：它只不过是各变元的幂的乘积的一个线性组合．如果诸变元是非交换的，那么相同变元的不同幂可以与其他变元的幂的积相间地出现在某个乘积之中．

2.4.15 定理 设 A，$B\in M_n$ 分别有 $\sigma(A)=\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ 和 $\sigma(B)=\{\beta_1, \beta_2, \cdots, \beta_n\}$(计算重特征值)．那么，存在非奇异矩阵 $S\in M_n$，使得 $S^{-1}AS$ 和 $S^{-1}BS$ 都是上三角矩阵，当且仅当存在指标 1，2，…，n 的一个排列 i_1，i_2，…，i_n，使得 $\sigma(p(A, B))=\{p(\alpha_j, \beta_{i_j}): j=1, \cdots, n\}$对两个(非交换)变元的所有复系数多项式 $p(t, s)$成立．

练习 验证，如果 A 和 B 可同时三角化，(2.4.15)中的多项式条件是必要的．证明，如

果 A，$B\in M_n$ 可交换，那么 $\sigma(p(A,B))=\{p(\alpha_j,\beta_{i_j})：j=1,\cdots,n\}$，对两个变元的所有多项式 p 成立. 定理(2.4.15)是如何适用于例(2.4.14)的?

[94]

附注 定理(2.4.15)的结论对任意域上的矩阵和多项式都能成立，只要这个域包含这两个矩阵的各特征值；对于 $k=3,4,\cdots$ 个矩阵的同时三角化，定理也成立(这时，定理的条件要涉及 k 个变化的多项式)；并且，就是对于特征值的限定子集，定理也成立，即 $p(\alpha_j,\beta_{i_j})\in\sigma(p(A,B))$，$j=1,\cdots,r$ 对所有多项式 $p(s,t)$ 成立，当且仅当 A 和 B 同时各自相似于这样一个分块三角矩阵，在这两个分块三角矩阵的对角线上的某些相应位置分别有由 $\alpha_1,\cdots,\alpha_r$ 和 $\beta_{i_1},\cdots,\beta_{i_r}$ 组成的 1×1 子块.

习题

1. 假定 A，$B\in M_n$ 可交换，且分别有特征值 $\alpha_1,\cdots,\alpha_n$ 和 $\beta_1,\cdots,\beta_n$. (a)证明，AB 的特征值是 $\alpha_1\beta_{i_1}$，$\alpha_2\beta_{i_2}$，$\cdots$，$\alpha_n\beta_{i_n}$，其中 $i_1,\cdots,i_n$ 是指标 $1,\cdots,n$ 的某个排列. (b)如果 $p(s,t)$是两个变元的多项式，证明 $p(A,B)$有特征值 $p(\alpha_1,\beta_{i_1}),\cdots,p(\alpha_n,\beta_{i_n})$. (c)最后证明，可同时上三角化这个较弱的假设足以得出上述结论：交换性不是必要的.

2. 如果 $A\in M_n$，证明 A 的秩不小于 A 的非零特征值的个数. **提示**：试证一个上三角矩阵的秩至少等于非零主对角元的个数，然后利用 Schur 定理(2.3.1). 用例子

$$A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$$

说明 A 的秩为什么可能大于非零特征值的个数.

3. 本题的目的是要证明，Cayley-Hamilton 定理对其元素取自交换环而不一定取自复数域的矩阵成立. 交换环是这样一个代数结构，除了乘法逆元的存在性以外，它满足域的所有公理. 因此其“加法”与“乘法”运算是交换的，且满足通常的结合律和分配律. 我们也明确假定在该环中存在乘法单位元；即存在元素“1”使得对该环中的所有 a 都有 $1a=a$. 是一个环而不一定是域的例子是 Z_k—所有整数对 k 的模. 在 Z_k 中，“加法”与“乘法”不是通常的运算，但其运算结果是对 k 取模；Z_k 是一个域当且仅当 k 是素数. 另一个例子是具有复系数的 k 个形式未定元的多项式集合.

[95]

(a) 我们知道，若 $A\in M_n$，则 adj $A\in M_n$ 是其 i，j 元为 A 的 j，i 代数余子式的唯一矩阵(见 0.8.2). 证明，基本公式

$$A(\text{adj }A)=(\text{adj }A)A=(\det A)I$$

正好是用代数余子式表示的 A 的行列式的 Laplace 展开式的一个表达式，并且证明，若 A 的任意两行或两列相等，则 $\det A=0$. 注意到这个公式只涉及乘法和加法而不涉及除法. 证明该公式对其元素取自任一交换环的矩阵是成立的.

(b) 利用(a)证明

$$(tI-A)[\text{adj}(tI-A)]=[\text{adj}(tI-A)](tI-A)=\det(tI-A)I=p_A(t)I$$

对任意 $A\in M_n$，甚至对其元素取自一个交换环 $n\times n$ 矩阵都成立. 证明，矩阵 adj$(tI-A)$是这样一个矩阵，其元素是次数不超过 $n-1$ 的 t 的多项式，因而它可以写成

$$\text{adj}(tI-A)=A_{n-1}t^{n-1}+A_{n-2}t^{n-2}+\cdots+A_1t+A_0,$$

其中系数 A_k 是 $n\times n$ 矩阵，其元素是 A 的诸元素的多项式函数. 多项式 $p_A(t)$ 是 A 的特征多项式.

(c) 证明，若 A 是一个其元素取自一个交换环的 $n\times n$ 矩阵，则

$$\begin{aligned} t^kI-A^k &= (tI-A)(It^{k-1}+At^{k-2}+\cdots+A^{k-2}t+A^{k-1}) \\ &\equiv (tI-A)G_k(A,t) \end{aligned}$$

对 $k=0$, 1, 2, …都成立. 由此得出

$$t^kI = It^k = A^k+(tI-A)G_k(A,t),\quad k=0,1,2,\cdots.$$

(d) 设 $p_A(t)=a_nt^n+a_{n-1}t^{n-1}+\cdots+a_1t+a_0=\det(tI-A)$ 是 A 的特征多项式(其中 $a_n=1$)，并且注意到它对其元素取自一个交换环的 $n\times n$ 矩阵也是有定义的. 利用(c)证明

$$\begin{aligned} p_A(t)I &= \sum_{k=0}^{n}a_kt^kI = \sum_{k=0}^{n}a_k[A^k+(tI-A)G_k(A,t)] \\ &= p_A(A)+(tI-A)G(A,t), \end{aligned}$$

其中

$$G(A,t)=\sum_{k=0}^{n}a_kG_k(A,t)$$

96

是一个次数不超过 $n-1$ 的 t 的多项式，以矩阵为系数，其元素是 A 的诸元素的多项式函数. 现在用(b)来证明

$$\begin{aligned} p_A(A) &= p_A(t)I-(tI-A)G(A,t) \\ &= (tI-A)\mathrm{adj}(tI-A)-(tI-A)G(A,t) \\ &= (tI-A)H(A,t)\equiv Q_A(t), \end{aligned}$$

其中 $H(A,t)=B_{n-1}t^{n-1}+B_{n-2}t^{n-2}+\cdots+B_1t+B_0$，而每个 B_k 是一个 $n\times n$ 矩阵，其元素是与 t 无关的 A 的诸元素的多项式函数. 因此 $Q_A(t)$ 是一个具有矩阵系数的次数不超过 n 的 t 的多项式.

(e) 计算 $Q_A(A)$，由此得出 $p_A(A)=0$.

4. 设 $A\in M_n$ 是非奇异方阵. 证明任何与 A 可交换的矩阵也与 A^{-1} 可交换. **提示**：利用(2.4.4)，然后直接验证.

5. 用(2.3.1)证明，如果 $A\in M_n$ 有特征值 λ_1, λ_2, …, λ_n，那么

$$\sum_{i=1}^{n}\lambda_i^k=\mathrm{tr}\,A^k,\quad k=1,2,\cdots$$

6. 证明，对

$$A=\begin{bmatrix}1&0&0\\0&2&0\\0&0&3\end{bmatrix}\quad 和\quad B=\begin{bmatrix}-2&1&2\\-1&-2&-1\\1&1&1\end{bmatrix},$$

以及对所有纯量 a, $b\in\mathbf{C}$, 有 $\sigma(aA+bB)=\{a-2b,\ 2a-2b,\ 3a+b\}$，但是 A 和 B 不能同时相似于上三角矩阵.

7. 利用(2.3)节习题 6 中的论据证明例(2.4.14)中的两个矩阵不可能同时上三角化. 试对习题 6 中两个矩阵作同样的证明.

8. 在证明两个矩阵不酉等价时，McCoy 定理(2.4.15)的指导思想可能有用的. 设 $p(t, s)$ 是两个非交换变元的复杂数多项式，又设 $A, B\in M_n$ 酉等价，即对某个酉矩阵 $U\in M_n$，有 $A=UBU^*$. 证明 $p(A, A^*)=UP(B, B^*)U^*$. 因此，如果 A 和 B 酉等价，则对每个具有两个非交换变元的复多项式 $p(t, s)$，$\mathrm{tr}\ p(A, A^*)=\mathrm{tr}\ p(B, B^*)$. 这与定理(2.2.6)有什么关系？

9. 设 $A\in M_n$，$B\in M_m$ 是给定的，且假定 A 和 B 没有公共特征值；即 $\sigma(A)\cap\sigma(B)$ 是空集. 利用 Cayley Hamilton 定理(2.4.2)证明方程 $AX-XB=0$，$X\in M_{n,m}$ 只有解 $X=0$. 由此得出结论，对每个已知 $C\in M_{n,m}$，方程 $AM-XB=C$ 有唯一解 $X\in M_{n,m}$. **提示**：如果 $AX=XB$，用归纳法证明，对所有 $k=1, 2, \cdots$，$A^kX=XB^k$，从而对任意多项式 $p(t)$，有 $p(A)X=Xp(B)$. 选取 $p(t)$ 为 A 的特征多项式便得到 $p_A(A)X=0=Xp_A(B)$. 因为 $p_A(B)=(B-\lambda_1 I)\cdots(B-\lambda_n I)\cdots(B-\lambda_n I)$，其中 $\lambda_1\cdots, \lambda_n$ 是 A 的特征多项式，所以矩阵 $p_A(B)$ 是非奇的，因而 $Xp_A(B)=0$ 只有零解 $X=0$. 由齐次方程解的唯一性，并且将(0.5k 和 l)应用于 $M_{n,m}$ 上的线性变换 $X\to T(X)=AX-BX$ 便可推出，对方程右边任一已知矩阵 C，$AX-XB=C$ 的解存在.

10. 利用习题 9 给出定理(2.4.8)的一个证明，要求递推步骤不超过 $k-1$ 步. **提示**：把 A 写成

$$A=\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ 0 & A_{22} & \cdots & A_{2k} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & A_{kk} \end{bmatrix}=\begin{bmatrix} A_{11} & R_1 \\ 0 & T \end{bmatrix},$$

其中每个 A_{ii} 是其主对角线上只有 λ_i 的上三角矩阵，且 $R_1=[A_{12}\cdots A_{1k}]$. 考虑

$$S=\begin{bmatrix} I & X \\ 0 & I \end{bmatrix},\quad S^{-1}=\begin{bmatrix} I & -X \\ 0 & I \end{bmatrix},$$

其中 X 与 R_1 同阶. 证明

$$S^{-1}AS=\begin{bmatrix} A_{11} & 0 \\ 0 & T \end{bmatrix}$$

只需选 X 适合 $A_{11}X-XT=-R_1$. 依次按行做下去便可得出 A 相似于 $\mathrm{diag}(A_{11}, A_{22}, \cdots, A_{kk})$.

11. A、$B\in M_n$ 已知，且考虑换位子 $C=AB-BA$. 证明 $\mathrm{tr}\ C=0$. 考察 $A=\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ 和 $B=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$，说明换位子不必是幂零矩阵；也就是说，即使一个换位子的各特征值的和肯定为零，它的某些特征值也可以是非零的.

12. 设 $A, B\in M_n$，$C=AB-BA$，且假定 A 与 C 可交换. 证明 C 必定是幂零矩阵. 试对习题 11 中的例子作出说明. **提示**：为什么存在非奇异矩阵 $S\in M_n$，使得 $SCS^{-1}=\mathrm{diag}(C_{11}, C_{22}, \cdots, C_{kk})\equiv C_1$，其中每个 $C_{ii}\in M_{n_i}$ 是上三角矩阵，$n_1+n_2+\cdots+n_k=n$，$\sigma(C_{ii})=\{\lambda_i\}$，$(i=1, 2, \cdots, k)$，且 $\lambda_i\neq\lambda_j$，$(i\neq j)$？设 $A_1\equiv SAS^{-1}$，$B_1\equiv SBS^{-1}$，且把 $A_1=(A_{ij})$ 和 $B_1=(B_{ij})$ 写成与 C_1 的分块对角形式同型的分块形式. 证明 $A_1C_1=C_1A_1$，然后用习题 9 证明，只

要 $k>1$ 且 $i\neq j$，就有 $A_{ij}=0$. 于是每个 $C_{ii}=A_{ii}B_{ii}-B_{ii}A_{ii}$ 有 $\operatorname{tr} C_{ii}=0$，因而 $\lambda_i=0$；而 $k=1$ 时；C 显然是幂零矩阵.

13. 采用习题 9 的记号. 利用定理(2.4.9)给出如下事实的另一个证明：若 A 与 B 没有公共特征值，则对于每个 $C\in M_n$，方程 $AX-XB=C$ 有唯一解. **提示**：考虑用 $T_1(X)=AX$，$T_2(X)=XB$ 定义的线性变换 T_1，T_2：$M_{n,m}\to M_{n,m}$. 证明 T_1 与 T_2 可交换，由(2.4.9)推出 T 的诸特征值是 T_1 和 T_2 的诸特征值之差. 证明，λ 是 T_1 的一个特征值当且仅当存在非零 $X\in M_{n,m}$ 使得 $AX=\lambda X=0$，这可以成立当且仅当 λ 是 A 的一个特征值[考虑 X 的非零列]. 因此 T_1 的特征值的集合与 A 相同，同样，T_2 的特征值的集合与 B 相同. 因而，若 A 与 B 没有公共特征值，则 T 是非奇异的. 若 x 是 A 的相应于特征值 λ 的一个特征向量，而 y 是 B^T 的相应于特征值 μ 的一个特征向量，考虑 $X=xy^T$，证明 $T(X)=(\lambda-\mu)X$，由此得出 T 的特征值的集合由 A 的特征值与 B 的特征值的所有可能差组成.

14. 设 $\mathscr{F}=\{A_i: i\in\mathscr{I}\}\subset M_n$ 是一个交换族. 证明，$\mathscr{F}$ 在下述意义下可以同时上三角化：其中任意一个给定的成员化简为(2.4.8)中的特殊形式，而其他成员化简成共形分块对角上三角形式. 即对于每个给定的 $A_0\in\mathscr{F}$，证明，存在一个非奇异的 $S\in M_n$，使得对所有的 $i\in\mathscr{I}$，$A_i=S\operatorname{diag}(T_1^{(i)},\cdots,T_k^{(i)})S^{-1}$，其中，对所有 $j=1,2,\cdots,k$，$n_1+n_2+\cdots+n_k=n$，以及所有 $i\in\mathscr{I}$，每个 $T_j^{(i)}\in M_{n_j}$ 是上三角矩阵，而每个 $T_j^{(0)}$ 的所有主对角元为 λ_j，若 $j\neq i$，还有 $\lambda_j\neq\lambda_i$. **提示**：选定 S 使得 $S^{-1}A_0S$ 有(2.4.8)中的特殊分块对角上三角形式. 注意矩阵族 $\{S^{-1}A_iS: i\in\mathscr{I}\}$ 是交换的. 把每个 $S^{-1}A_iS$ 分成与 $S^{-1}A_0S$ 的分块形式共形的分块矩阵，然后利用交换性以及习题 9 或 13 的结果(像习题 12 中那样)，证明每个 $S^{-1}A_iS$ 的所有非对角子块一定化为零. 现在可以在处于相应对角子块的 k 个族上应用定理(2.3.3). 除了 $S^{-1}A_0S$ 以外，当然不保证 $S^{-1}A_iS$ 的一个对角子块的特征值都相等或不同的对角子块有不相交的谱.

进一步阅读和注释 定理(2.4.15)及其推广是由 N. H. McCoy 证明的，可参看 N. H. McCoy, "On the Characteristic Roots of Matrix Polynomials," *Bull. Amer. Math. Soc.* 42 (1936), 592-600. 也可参看 T. S. Motzkin and O. Taussky, "Pairs of Matrices with Property L," *Trans. Amer. Math. Soc.* 73(1952), 108-114，其中讨论了特征值与线性组合的关系. 一对 A，$B\in M_n$ 有性质 L，是指对所有 a，$b\in\mathbf{C}$ 有 $\sigma(aA+bB)=\{a\alpha_j+b\beta_j: j=1,\cdots,n\}$，而定理(2.4.15)的条件称为性质 P. 显然性质 P 蕴涵性质 L，反之不成立. 较弱的性质 L 尚未彻底弄清楚，但还是知道一些，例如，一对具有性质 L 的正规矩阵[见(2.5)节]一定可交换，因而一定可同时酉对角化.

2.5 正规矩阵

正规矩阵是在讨论酉等价时自然产生的一类矩阵，它在整个矩阵分析中是很重要的，并且它还推广了酉矩阵，实对称矩阵和 Hermite 矩阵.

2.5.1 定义 设矩阵 $A\in M_n$，如果 $A^*A=AA^*$，即，如果 A 与它的 Hermite 伴随可交换，就称 A 为正规矩阵.

练习 证明，$A\in M_n$ 是正规矩阵，当且仅当酉等价于 A 的每一个矩阵都是正规矩阵. 这

99

说明正规矩阵类在酉等价下是封闭的.

2.5.2 例

(a) 如果 U 是酉矩阵，那么 $U^*U=I=UU^*$，因而，所有酉矩阵是正规矩阵.

(b) 如果 $A^*=A$，显然 $A^*A=AA^*$，因而，所有 Hermite 矩阵都是正规矩阵.

(c) 如果 $A\in M_n$ 适合 $A^*=-A$，就称 A 是斜 Hermite 矩阵. 这时，$A^*A=-A^2=AA^*$，因而所有斜 Hermite 矩阵也都是正规矩阵.

(d) $A=\begin{bmatrix}1 & -1\\1 & 1\end{bmatrix}$是正规矩阵，但它不属于上述任一类矩阵.

练习 用元素间的关系给出 $M_2(\mathbf{R})$中的正规矩阵的特征. 并且按(2.5.2a，b 和 c)的分类作出回答. **提示：**如果 $A\in M_2(\mathbf{R})$是正规矩阵，证明，如果 A 至少有一个元素为零，那么或 $A=A^T$,或 $A=-A$. 如果 A 的所有元素是非零的，证明，$A-A^T$，或对某个 $a>0$，$AA^T=aI$.

100

练习 给出一个 2×2 非正规实矩阵的例子. 同时给出一个 2×2 实矩阵的例子，它是正规矩阵，但不是对称的，斜对称的或正交的.

练习 证明(2.5.2a，b 和 c)的每一类各自在酉等价下是封闭的.

练习 证明，Hermite 对角矩阵必定是实的，而斜 Hermite 对角矩阵必定是纯虚的.

2.5.3 定义 如果 $A\in M_n$ 酉等价于一个对角矩阵，就称 A 是可酉对角化，类似地可定义可正交对角化概念. 应注意的是，"可酉(或正交)对角化"蕴涵可对角化(但反之不成立).

练习 考察(1.3.7)的证明便可得出，$A\in M_n$ 可酉对角化，当且仅当在 $\mathbf{C}^n$ 中有一个由 n 个标准正交向量组成的集合，其中每个向量都是 A 的特征向量.

下面，列出关于正规矩阵的最基本的事实. 在下述定理中，(a)与(b)的等价性常常被称为正规矩阵的谱定理.

2.5.4 定理 如果 $A=[a_{ij}]\in M_n$ 有特征值 $\lambda_1,\cdots,\lambda_n$，那么下述命题等价：

(a) A 是正规矩阵；

(b) A 是可酉对角化矩阵；

(c) $\sum\limits_{i,j=1}^{n}|a_{ij}|^2=\sum\limits_{i=1}^{n}|\lambda_i|^2$；

(d) 存在由 A 的 n 个特征向量组成的标准正交基.

证明：由 Schur 定理保证，在该证明中，始终假定 $T=[t_{ij}]\in M_n$ 是酉等价于 A 的上三角矩阵；即对某个酉矩阵 $U\in M_n$，有 $T=U^*AU$，因为 T 酉等价于 A，所以命题(a)等价于 T 的正规性. 我们将证明，(a)与(b)等价，(b)与(c)等价，(b)与(d)等价，从而完成证明.

为了证明(a)蕴涵(b)，需做一些计算. 如果 A 是正规矩阵，那么 T 也是正规矩阵. 但一个正规的三角矩阵必定是对角矩阵. 这只要令 T^*T 和 TT^* 的对角元相等就可以看出来. T^*T 的 1，1 元与 TT^* 的 1，1 元相等表明

101

$$\bar{t}_{11}t_{11}=t_{11}\bar{t}_{11}+\sum_{j=2}^{n}t_{1j}\bar{t}_{1j}=|t_{11}|^2+\sum_{j=2}^{n}|t_{1j}|^2.$$

这就是说，$0 = \sum_{j=2}^{n} |t_{1j}|^2$，若干非负项的和为零，那么它的每一项必须为 0. 由此得出

$$t_{1j} = 0, \quad j = 2, \cdots, n.$$

T^*T 和 TT^* 的 2，2 元相等表明

$$\bar{t}_{22}t_{22} = t_{22}\bar{t}_{22} + \sum_{j=3}^{n} t_{2j}\bar{t}_{2j} = |t_{22}|^2 + \sum_{j=3}^{n} |t_{2j}|^2,$$

同理可得

$$t_{2j} = 0, \quad j = 3, \cdots, n.$$

假定采用同样的方法已经验证

$$t_{ij} = 0, \quad j > i \text{ 和 } i = 1, 2, \cdots, k-1$$

由此可以得出

$$t_{ij} = 0, \quad j > i, \quad i = k,$$

依次讨论 T^*T 和 TT^* 的每个对角元，最后得到

$$t_{ij} = 0, \quad j > i, \quad i = 1, 2, \cdots, n,$$

因为 T 是上三角矩阵，所以又有

$$t_{ij} = 0, \quad j < i, \quad i = 1, 2, \cdots, n.$$

因此，证明了 T 是对角矩阵，因而(b)成立. 由于对角矩阵显然是正规矩阵，又酉等价保持正规性，所以(b)也蕴涵(a).

要证明(b)与(c)等价，需借助于(2.2.2). 因为任一与 A 酉等价的对角矩阵的对角元是特征值 λ_1，…，λ_n(按某个顺序排列)，(2.2.2)使我们从(b)推出(c). 另一方面，因为 λ_i，$i=1$，…，n，是 T 的对角元(按某个顺序排列)，所以，(2.2.2)也表示

$$\sum_{i,j=1}^{n} |a_{ij}|^2 = \sum_{i=1}^{n} |\lambda_i|^2 + \sum_{i<j} |t_{ij}|^2.$$

但是(c)意味着

$$\sum_{i<j}^{n} |t_{ij}|^2 = 0.$$

或 T 是对角矩阵. 由此推出(b)成立. [102]

(b)与(d)的等价性，是本定理之前的那个练习所要证明的事实. []

练习 如果 $T \in M_n$ 是三角矩阵，且 T^*T 与 TT^* 的第 i 个对角元相同，$i=1$，…，n，证明 T 是对角矩阵. 说明为什么这个结论连同正规性在酉相似下是不变的结论是正规矩阵可酉对角化的基本理由.

练习 证明一个正规矩阵是非亏损的(每个特征值几何重数等于它的代数重数).

练习 证明，如果 $A \in M_n$ 是正规矩阵，那么 $x \in \mathbf{C}^n$ 是 A 的相应于特征值 λ 的右特征向量，当且仅当 x 是 A 的相应于 λ 的左特征向量；即 $Ax = \lambda x$ 等价于 $x^*A = \lambda x^*$. **提示**：正规化 x，且记 $A = U\Lambda U^*$，使 x 为 U 的第 1 列. 那么，A^* 及 A^*x 各是什么？参看本节后面的习题 20 中的另一个证明.

练习 如果 $A \in M_n$ 是正规矩阵，且 x 和 y 是相应于不同特征值的特征向量，证明 x 与 y

正交. **提示**: 如果 $Ax=\lambda x$, $Ay=\mu y$, 试证 $\mu x^* y=x^*(Ay)=(A^* x)^* y=(\bar{\lambda}x)^* y=\lambda x^* y$. 如果 $\lambda\neq\mu$, 由此得出 $x^* y=0$. 参看习题21的另一个证明.

如果已知一个正规矩阵的各特征值, 那么它可以通过下面理论描述酉对角化. 确定每个特征空间, 然后求它的一个标准正交基(例如, 用 Gram-Schmidt 过程). 因为每个特征空间的维数等于其相应特征值的重数, 又 A 是正规矩阵, 所以这些基的并就是整个空间的一个标准正交基. 把这些向量排成一个酉矩阵的列便得到所要求的对角化变换.

下面说明, 可交换的正规矩阵可同酉对角化.

2.5.5 定理 如果 $\mathcal{A}\subseteq M_n$ 是正规矩阵的交换族, 那么 $\mathcal{A}$ 可同时酉对角化: 即存在一个酉相似, 它把 $\mathcal{A}$ 中每个矩阵变换成对角矩阵.

练习 利用(2.3.3)以及正规的三角矩阵一定是对角矩阵的事实证明(2.5.5). 说明(2.5.5)的条件和结论要比(1.3.19)的强.

[103]

将(2.5.4)应用于 Hermite 矩阵的特殊情形, 便得到一个基本结果, 常常称之为 Hermite 矩阵的谱定理.

2.5.6 定理 如果 $A\in M_n$ 是 Hermite 矩阵, 那么

(a) A 的所有特征值都是实的;

(b) A 是可酉对角化的.

如果 $A\in M_n(\mathbf{R})$ 是对称矩阵, 那么 A 可实正交对角化.

证明: 一个 Hermite 对角矩阵必定具有实对角元, 因而(a)可以从(b)和 Hermite 矩阵的集合在酉等价下封闭的事实推出. 因为 Hermite 矩阵是正规矩阵, 所以命题(b)可以从(2.5.4)推出. 如果 $A\in M_n(\mathbf{R})$ 是对称矩阵, 那么它就是 Hermite 矩阵, 不过, 我们注意到, 对角化 A 所需一切计算均可在实数域上进行. 因为 A 的特征值是实数, 所以相应的特征向量可以取实向量. □

在(2.5.4)和(2.5.6)中, 特征值相同或不同已无关紧要了. 在(2.5.5)中不必假定可对角化条件. 将这些与第1章中可对角化的讨论作一比较是有意义的. Hermite 矩阵和正规矩阵从结构上就保证它们有一个线性无关的完备的特征向量组(实际上是标准正交组), 这就是为什么它们如此重要且有如此合意的性质的理由.

作为结束, 我们证明与(2.5.4)和(2.5.5)相类似的定理, 它们是关于实正规矩阵的. 这些矩阵是正规矩阵, 因此可经一个未必是实的酉相似对角化. 如果只经一个实正交相似, 那么这样的实正规矩阵可能变成的最好形式是怎样的呢? 因为一个实正规矩阵可能没有任何实的特征值, 所以不能保证它可经一个实相似对角化. 另一方面, 任一实矩阵可经一个实正交相似变成特殊的分块上三角形式(见2.3.4), 如果这个矩阵还是正规矩阵, 就假定它已变成这种特殊的形式. 证明要利用(2.3.4), 正像在(2.5.4)中用到了(2.3.1)一样. 下面的引理可消除一个小的技术障碍, 它在(2.5.4)的证明中不曾出现.

2.5.7 引理 如果 $A\in M_n$ 是 Hermite 矩阵, 且对所有 $x\in\mathbf{C}^n$, 有 $x^* Ax\geqslant 0$, 那么 A 的所有特征值非负. 如果还有 $\operatorname{tr} A=0$, 那么 $A=0$.

证明：利用(2.5.6)可以写出 $A=U\Lambda U^*$，其中，$U=[u_1u_2\cdots u_n]\in M_n$ 是酉矩阵，而 $\Lambda=\operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$．于是 $\Lambda=U^*AU$，根据假设可知，$\lambda_k=u_k^*Au_k\geqslant 0$，从而所有 $\lambda_k\geqslant 0$．最后，$\operatorname{tr}A=\operatorname{tr}U\Lambda U^*=\operatorname{tr}\Lambda U^*U=\operatorname{tr}\Lambda=\lambda_1+\cdots+\lambda_n$，因此，如果 $\operatorname{tr}A=0$ 和所有 $\lambda_k\geqslant 0$，就一定有所有 $\lambda_k=0$，因而 $\Lambda=0$ 且 $A=U\Lambda U^*=U0U^*=0$． □ 104

2.5.8 定理 设 $A\in M_n(\mathbf{R})$，则 A 是正规矩阵，当且仅当存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$，使得

$$Q^tAQ=\begin{bmatrix}A_1 & & & 0\\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_k\end{bmatrix}\in M_n(\mathbf{R}),\quad 1\leqslant k\leqslant n, \tag{2.5.9}$$

其中每个 A_i 或者是 1×1 矩阵，或者是形如

$$A_j=\begin{bmatrix}\alpha_j & \beta_j\\ -\beta_j & \alpha_j\end{bmatrix} \tag{2.5.10}$$

的 2×2 矩阵．

证明：由直接计算可知，每个形如(2.5.10)的矩阵是正规矩阵$[A_jA_j^T=\operatorname{diag}(\alpha_j^2+\beta_j^2, \alpha_j^2+\beta_j^2)=A_j^TA_j]$，因而任一个形如(2.5.9)的直和也一定是正规矩阵．由于前面的论断，如果利用(2.3.4)来证明定理，那只需对形如(2.3.5)的实正规矩阵来证明就够了．因为在(2.3.5)中诸主对角子块可以按任意规定的顺序排列，所以可以假定

$$A=\begin{bmatrix}R & A_{01} & A_{02} & \cdots & A_{0k}\\ & A_{11} & A_{12} & \cdots & A_{1k}\\ & & A_{22} & \cdots & A_{2k}\\ & 0 & & \ddots & \vdots\\ & & & & A_{kk}\end{bmatrix}\in M_n(\mathbf{R}) \tag{2.5.11}$$

是正规矩阵，其中

$$R=\begin{bmatrix}\lambda_1 & & *\\ & \ddots & \\ 0 & & \lambda_p\end{bmatrix}\in M_p(\mathbf{R})$$

是上三角矩阵．$A_{01}, A_{02}, \cdots, A_{0k}\in M_{p,2}(\mathbf{R})$，并且对 $i, j=1, 2, \cdots, k$ 以及 $j\geqslant i$，有 $A_{ij}\in M_2(\mathbf{R})$．下面要证明 R 是对角矩阵，且对所有 $j>i$，$A_{ij}=0$．

比较恒等式 $A^TA=AA^T$ 中与(2.5.11)中的子块 R 相对应的第一个 $p\times p$ 主对角块元，得到恒等式

$$R^TR=RR^T+A_{01}A_{01}^T+\cdots+A_{0k}A_{0k}^T. \tag{2.5.12}$$ 105

我们知道，对某个 $E\in M_{p,q}$，每个形如 $B=EE^*$ 的矩阵 $B\in M_p(\mathbf{C})$是 Hermite 矩阵，且对所有 $x\in\mathbf{C}^n$有性质 $x^*Bx=x^*EE^*x=(E^*x)^*(E^*x)\geqslant 0$，并且，这样一些矩阵的和也有同样的性

质. 因为，按一般的原则，

$$\operatorname{tr} R^T R = \operatorname{tr} RR^T,$$

而且，由(2.5.12)有

$$\operatorname{tr} R^T R = \operatorname{tr} RR^T + \operatorname{tr} A_{01}A_{01}^T + \cdots + \operatorname{tr} A_{0k}A_{0k}^T,$$

由此可见

$$0 = \operatorname{tr} A_{01}A_{01}^T + \cdots + \operatorname{tr} A_{0k}A_{0k}^T.$$

根据引理(2.5.7)，且把上述结论应用于实矩阵 $B=A_{0j}A_{0j}^*=A_{0j}A_{0j}^T$，有 $\operatorname{tr} A_{0j}A_{0j}^T \geqslant 0$. 因为它们的和是零，所以每一项也是零，因而 $A_{0j}A_{0j}^T=0$，$j=1$，…，k. 因为 $A_{0j}A_{0j}^T$ 的第 i 个主对角元是 A_{0j} 的第 i 行的(实)元素的平方和，因此，这些元素必须全为零，从而 $A_{0j}=0$，$j=1$，2，…，k，且(2.5.12)简化成

$$R^T R = RR^T.$$

但是，已经在(2.5.4)中证明，一个正规的三角矩阵必定是对角矩阵，所以有 $R=\operatorname{diag}(\lambda_1, \cdots, \lambda_p)$，这正是我们所断言的.

比较恒等式 $A^TA=AA^T$ 中与(2.5.11)中子块 A_{11} 相对应的 2×2 主对角块元，利用对所有 $j=1$，2…，k 有 $A_{0j}=0$ 的事实，便得到恒等式

$$A_{11}^T A_{11} = A_{11}A_{11}^T + A_{12}A_{12}^T + \cdots + A_{1k}A_{1k}^T. \tag{2.5.13}$$

而 $\operatorname{tr}(A_{11}^TA_{11})$，像上面那样运用引理(2.5.7)，于是，对 $j=2$，3…，k，有

$$\operatorname{tr}(A_{12}A_{12}^T) + \cdots + \operatorname{tr}(A_{1k}A_{1k}^T) = 0, \quad \operatorname{tr}(A_{1j}A_{1j}^T) \geqslant 0,$$

$$\operatorname{tr}(A_{1j}A_{1j}^T) = 0, \quad A_{1j}A_{1j}^T = 0 \text{ 和 } A_{1j} = 0,$$

因此(2.5.13)简化成 $A_{11}^TA_{11}=A_{11}A_{11}^T$；即 2×2 子块 A_{11} 是正规矩阵.

依次验证恒等式 $A^TA=AA^T$ 的与(2.5.11)中子块 A_{ii} 相对应的 2×2 主对角块元，其中 $i=2$，3，…，$k-1$，并且运用同样的论证，如我们所断言的那样，可以得出，所有非对角子块都是零，且所有对角子块 A_{ii} 都是正规矩阵.

我们已经证明，一个实正交相似可把一个实正规矩阵化成形式(2.3.5). 实际上可化成分块对角形式(2.5.9). 只要证明所有对角子块有形式(2.5.10)，那就完成了整个证明.

106

如果 $A_{jj}=\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2(\mathbf{R})$ 是正规矩阵，那么，使恒等式 $A_{jj}^TA_{jj}=A_{jj}A_{jj}^T$ 的 1，1 元和 1，2 元分别相等便得到恒等式

$$b^2 = c^2, \quad \text{从而 } c = \pm b,$$

以及

$$ac+bd=ab+cd, \text{ 从而，如果 } c=-b, \text{ 则}$$

$$2b(a-d) = 0.$$

$c=+b$ 及 $b=0$ 的情形可以排除在外，因为这时 A_{jj} 将是实对称矩阵，因而只有实特征值；但是按我们的作法，子块 A_{jj} 有一对共轭的非实特征值. 因此，只能有 $c=-b$，$a=d$，且 A_{jj} 必有形式(2.5.10). 由计算可知，实矩阵 $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ 有一对共轭的复特征值 $\lambda=a+a+ib$ 和 $=\lambda a-ib$.

□

作为这个关于实正规矩阵的定理的推论，容易推导出实矩阵是对称矩阵、斜对称矩阵或正交矩阵时的实标准形.

2.5.14 推论 设 $A\in M_n(\mathbf{R})$. 于是有

(a) $A=A^T$，当且仅当存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$，使得

$$Q^TAQ=\begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix},\quad \text{其中所有 } \lambda_i\in\mathbf{R};$$

(b) $A=-A^T$，当且仅当存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$，使得

$$Q^TAQ=\begin{bmatrix}0 & & & & & 0\\ & \ddots & & & & \\ & & 0 & & & \\ & & & A_1 & & \\ & 0 & & & \ddots & \\ & & & & & A_k\end{bmatrix},$$

其中每个 $A_j\in M_2(\mathbf{R})$ 有形式

$$A_j=\begin{bmatrix}0 & \beta_j\\ -\beta_j & 0\end{bmatrix};$$

107

(c) $AA^T=I$，当且仅当存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$，使得

$$Q^TAQ=\begin{bmatrix}\lambda_1 & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_p & & & \\ & & & A_1 & & \\ & 0 & & & \ddots & \\ & & & & & A_k\end{bmatrix},$$

其中，每个 $\lambda_j=\pm1$，而每个 $A_j\in M_2(\mathbf{R})$ 有形式

$$A_j=\begin{bmatrix}\cos\theta_j & \sin\theta_j\\ -\sin\theta_j & \cos\theta_j\end{bmatrix},\quad \theta_j\in\mathbf{R}.$$

证明： 在每一种情形，定理的条件都保证 A 是实正规矩阵，因此 A 可写成形式(2.5.9)和(2.5.10). 如果 $A=A^T$，那么每个 $A_j=A_j^T$，因而所有 $\beta_i=0$，而 Q^TAQ 是对角矩阵. 如果 $A=-A^T$,那么每个 $\lambda_j=-\lambda_j$ 且每个 $A_j=-A_j^T$，因而所有 $\lambda_j=0$ 且所有 $\alpha_j=0$. 如果 $AA^T=I$，那么每个 $\lambda_j\lambda_j=1$ 且每个 $A_jA_j^T=I$，因而，所有 $\lambda_j^2=1$ 且所有 $\alpha_j^2+\beta_j^2=1$；这时有 $\lambda_j=\pm1$ 且 $\alpha_j=\cos\theta_j,\beta_j=\sin\theta_j$. □

如果有一个由可交换的实正规矩阵组成的族，它们可能不可同时实对角化，但是可以同时把它们都变成分块对角形式(2.5.9).

2.5.15 定理 如果 $\mathcal{F}\subseteq M_n(\mathbf{R})$ 是实正规矩阵的交换族，那么存在实正交矩阵 Q，使得对所有 $A\in\mathcal{F}$，Q^TAQ 具有形式(2.5.9)和(2.5.10).

证明： 运用(2.3.6)，可经实正交相似 Q 把 $\mathcal{F}$ 的每个成员化成形式(2.3.5). (2.3.8)的证明过程说明，它们这时有形式(2.5.9). □

习题

关于 $A\in M_n$ 的与正规性等价的条件，除了(2.5.4)所列出的以外，还可以罗列出许多. 在下面的习题中，其中就包含若干等价条件.

1. 证明，$A\in M_n$ 是正规矩阵，当且仅当对所有 $x\in\mathbf{C}^n$，Ax 与 A^*x 的 Euclid 长度相同. 我们知道，$(y^*y)^{1/2}$ 是 $y\in\mathbf{C}^n$ 的 Euclid 长度.

108

2. 证明，正规矩阵是酉矩阵，当且仅当它的所有特征值有绝对值 1.

3. 证明，正规矩阵是 Hermite 矩阵，当且仅当它的所有特征值都是实数.

4. 证明，正规矩阵是斜 Hermite 矩阵，当且仅当它的所有特征值都是纯虚数(有等于 0 的实部).

5. 如果 $A\in M_n$ 是斜 Hermite(相应地，Hermite)矩阵，证明 iA 是 Hermite(相应地，斜 Hermite)矩阵.

6. 证明，$A\in M_n$ 是正规矩阵，当且仅当它与一个具有互异特征值的正规矩阵可交换.

7. 如同定理(2.1.9)那样，假定 $A\in M_n$ 有形式 $A=B^{-1}B^*$，其中 $B\in M_n$ 是非奇异矩阵. (a)证明，A 是酉矩阵，当且仅当 B 是正规矩阵. (b)如果 B 有形式 $B=HNH$，其中，N 是正规矩阵，而 H 是 Hermite 矩阵(且它们都是非奇异矩阵)，证明 A 相似于一个酉矩阵.

8. 定义 $H(A)=\frac{1}{2}(A+A^*)$ 为 $A\in M_n$ 的 Hermite 部分，而 $S(A)=\frac{1}{2}(A-A^*)$ 为 A 的斜 Hermite 部分，那么 $A=H(A)+S(A)$. 证明 A 是正规矩阵，当且仅当 $H(A)$ 与 $S(A)$ 可交换.

9. 如果两个正规矩阵可交换，证明它们的乘积是正规矩阵，并用例子说明，即使两个正规矩阵不可交换，它们的乘积也可以是正规矩阵.

10. 采用习题 8 的记号，证明，如果 $H(A)$ 的每个特征向量是 $S(A)$ 的(相应地，A 的)特征向量，那么 A 是正规矩阵.

11. 对于任意复数 $z\in\mathbf{C}$，证明存在某个 $\theta\in\mathbf{R}$，使得 $\bar{z}=e^{i\theta}z$. 注意 $[e^{i\theta}]\in M_1$ 是酉矩阵. 想一想，酉对角矩阵 $U\in M_n$ 像什么呢?

12. 推广习题 11，证明，如果 $\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_n)\in M_n$，那么存在酉对角矩阵 U，使得 $\bar{\Lambda}=U\Lambda=\Lambda U$.

13. 利用习题 12 证明，矩阵 $A\in M_n$ 是正规矩阵当且仅当存在酉矩阵 $V\in M_n$，使得 $A^*=AV$. 这与习题 7 有什么关系?

14. 如果 $A\in M_n(\mathbf{R})$，且 A 的所有特征值都是实数，证明，A 是正规矩阵，当且仅当它是对称矩阵.

109

15. 证明，两个同阶正规矩阵相似(实际上是酉等价)，当且仅当它们有相同的特征多项式. 这对非正规矩阵也成立吗? **提示：** 考察 $\begin{bmatrix}0&0\\0&0\end{bmatrix}$ 和 $\begin{bmatrix}0&1\\0&0\end{bmatrix}$.

16. 如果 $A, B\in M_n$ 是正规矩阵，说明 AB 未必是正规矩阵．并说明某个阶数的非奇异正规矩阵不构成一个乘法群．但是酉正规矩阵确实构成一个群．非奇异 Hermite 矩阵构成一个乘法群吗？

17. 如果 $A\in M_n$ 是正规矩阵，且 $p(t)$是给定的多项式．用定义(2.5.1)证明 $p(A)$是正规矩阵．利用(2.5.4)对这个结果给出另一个证明．

18. 如果 $A\in M_n$ 有性质：对某个非零多项式 $p(t)$，$p(A)$是正规矩阵．那么 A 是正规矩阵吗？**提示**：考察 $A=\begin{bmatrix}0 & 1\\ -2 & 0\end{bmatrix}$ 和 A^2．

19. 设 $A\in M_n$，且 $\alpha\in \mathbf{C}$ 是给定的．证明，A 是正规矩阵，当且仅当 $A+\alpha I$ 是正规矩阵．

20. 设 $A\in M_n$ 是正规矩阵，并且假定 $x\in \mathbf{C}^n$ 是使 $Ax=\lambda x$ 的向量．试利用习题 1 和 19 证明$A^*x=\bar{\lambda}x$．**提示**：如果向量$(A-\lambda I)x$ 的 Euclid 长度是零，试证$(A-\lambda I)^*x$ 的 Euclid 长度也是零．

21. 利用(2.5.4)证明，如果 $A\in M_n$ 是正规矩阵，且 $Ax=\lambda x$ 和 $Ay=\mu y$，其中 $\lambda\neq\mu$，那么 x 与 y 正交．**提示**：设 $A=U\Lambda U^*$，其中 $U\in M_n$ 是酉矩阵，而 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$．令 $U^*x=x'=[x_i']$和 $U^*y=y'=[y_i']$．证明 $\Lambda x'=\lambda x'$，且由此推出对每个使 $\lambda_i\neq\lambda$ 的指标 i，有 $x_i'=0$，同理，y'也有类似性质．再证明 x'与 y'正交，从而得出 x 与 y 正交．

22. 利用(2.5.6)证明，即使 Hermite 矩阵 A 的所有元素不都是实数，它的特征多项式必定有实系数．

23. 证明，$\begin{bmatrix}1 & i\\ i & 1\end{bmatrix}$和$\begin{bmatrix}i & i\\ -i & 1\end{bmatrix}$都是复对称矩阵($A=A^T$)，但一个是正规矩阵，而另一个则不是．因此，在实对称矩阵和复对称矩阵之间存在着很大的差别[见(4.4)节]．

24. 如果 $A\in M_n$ 既是正规矩阵，又是幂零矩阵，证明 $A=0$．

25. 设 $A\in M_n$ 是给定的矩阵．证明，A 是正规矩阵，当且仅当存在次数至多为 $n-1$ 的多项式 $p(t)$，使得 $A^*=p(A)$．**提示**：如果 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，用 Lagrange 插值法构造一个多项式 $p(t)$使 $p(\Lambda)=\bar{\Lambda}$，然后引用(2.5.4)．这是如何"说明"为什么一个正规矩阵与它的伴随可交换的？此外，如果 A 是实矩阵，证明使得 $A^*=p(A)$的 Lagrange 插值多项式 $p(\cdot)$有实系数且 $A^T=p(A)$．因此对实正规矩阵 A 有 $A^T=p(A)$，其中 $p(\cdot)$是实多项式．见(0.9.11.4)．

26. 给出一个实正规矩阵的例子，它酉相似于一个对角矩阵，但不实正交相似于一个对角矩阵．证明，实矩阵 A 实正交相似于一个对角矩阵，当且仅当 A 是对称矩阵($A=A^T$)．

27. 证明，已知矩阵 $A\in M_n$ 是正规矩阵，当且仅当对所有 $x, y\in \mathbf{C}^n$ 有

$$(Ax)^*(Ay)=(A^*x)^*(A^*y).$$

从几何上看，这表示对所有 $x, y\in \mathbf{C}^n$，Ax 与 Ay 之间的夹角等于 A^*x 与 A^*y 之间的夹角．这与习题 1 是什么关系？

28. 如果 $A\in M_n$ 是正规矩阵，证明，$Ax=0$，当且仅当 $A^*x=0$．这表明 A 的零空间与 A^* 的零空间相同．考察$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$和$\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}$，说明这对一般情形不成立．

29. 假定有线性方程组 $Ax=y$，其中，$A\in M_n$ 和 $y\in \mathbf{C}^n$ 是已知的，且假定 A 是奇异矩阵．所给定的方程组有(非唯一的)解，当且仅当对每个使 $A^*z=0$ 的 $z\in \mathbf{C}^n$[见(0.6.6)]有式 $y^*z=0$．可是，如果 A 是正规矩阵，便可证明，所给定的方程组有解，当且仅当对每个使 $Aw=0$ 的 $w\in \mathbf{C}^n$ 有 $y^*w=0$；即 y 正交于 A 的零空间．如果想求奇异方程组 $Ax=y$ 的所有解，说明为什么在计算上 A 是正规矩阵时的求解比 A 不是正规矩阵时更简单．

30. 设 n_1，n_2，…，n_k 是给定的正整数，且 $A_j\in M_{n_j}$，$j=1$，2，…，k．证明，直和 $A=A_1\oplus A_2\oplus\cdots\oplus A_k$ 是正规矩阵，当且仅当每个 A_j 都是正规矩阵．

31. 证明，两个正规矩阵相似，当且仅当它们酉等价．**提示**：证明，如果 U 和 V 是酉矩阵，那么 UAU^* 和 VAV^* 酉等价．给出一个例子，说明两个(非正规)矩阵相似，但不酉等价．

32. 如果 $A\in M_3(\mathbf{R})$是实正交矩阵，我们知道，A 要么有一个，要么有三个实特征值．如果它有一个正的行列式，利用(2.5.14)证明它正交等价于$[1]\in M_1$ 与一个平面旋转的直和，从几何上讨论，可把这看作绕某条经 $\mathbf{R}^3$ 中原点的固定轴转动 θ 角的旋转．这是关于力学的 Euler

[111] 定理的一部分：每个刚体运动是一个平移与一个绕某条轴的旋转的合成．

33. 如果 $\mathscr{F}\subseteq M_n$ 是正规矩阵的交换族，证明存在 Hermiter 矩阵 B，使得对每个 $A_\alpha\in\mathscr{F}$，都有次数至多为 $n-1$ 的多项式 $p_\alpha(t)$使 $A_\alpha=p_\alpha(B)$．应当指出，B 对于 $\mathscr{F}$ 的所有元素是固定的，而多项式可能与 $\mathscr{F}$ 的元素有关．**提示**：设 $U\in M_n$ 是同时对角化 $\mathscr{F}$ 的每个成员的酉矩阵，设 $B=U\mathrm{diag}(1, 2, \cdots, n)U^*$，$A_\alpha=U\Lambda_\alpha U^*$，其中 $\Lambda_\alpha=\mathrm{diag}(\lambda_1^{(\alpha)}, \cdots, \lambda_n^{(\alpha)})$．然后取 $p_\alpha(t)$是使 $p_\alpha(k)=\lambda_k^{(\alpha)}$，$k=1$，2，…，$n$，的 Lagrange 插值多项式．

34. 证明，$A\in M_n$ 是正规矩阵，当且仅当 A 的每个特征向量也是 A^* 的一个特征向量．**提示**：设$U\in M_n$ 是酉矩阵，它的第 1 列是 A(因而是 A^*)的一特征向量．检验 U^*AU 和 $U^*A^*U=(U^*AU)^*$，然后继续做下去．

35. 当 A，$B\in M_n$ 是正规矩阵时，证明(2.2.8)的下述改进：A 酉等价于 B，当且仅当 $\mathrm{tr}\, A^k=\mathrm{tr}\, B^k$，$k=1$，2，…，$n$．**提示**：利用习题 15 以及(1，2)节习题 12．

36. 设 $A\in M_n(\mathbf{R})$，且假定 $AA^T=A^TA$，因而 A 是实正规矩阵．如果 AA^T 的特征值各不相同，证明 A 必须是对称矩阵．**提示**：利用定理(2.5.8)．

2.6 *QR* 分解和 *QR* 算法

为一个给定的矩阵 $A\in M_n$ 提供一种 Schur 酉上三角化(2.3.1)的特殊计算方法，以及(在某些假设下)计算特征值的一个通用的数值方法，称为 QR 算法，QR 算法基于一般矩阵 $A\in M_{n,m}$的所谓 QR 分解．

2.6.1 定理(QR 分解)　如果 $A\in M_{n,m}$ 且 $n\geqslant m$，那么存在有标准正交列的矩阵 $Q\in M_{n,m}$和上三角矩阵 $R\in M_m$，使得 $A=QR$．如果 $n=m$，那么 Q 是酉矩阵；此外，如果 A 是非奇异矩阵，则可以选取 R 为具有正对角元的上三角矩阵，并且在这种情形，因子 Q 和 R 都是唯一的．如果 $A\in M_{n,m}(\mathbf{R})$，那么 Q 和 R 都可以取实矩阵．

证明：如果 $A\in M_{n,m}$，且 $\mathrm{rank}\, A=m$，则 A 的各列构成 $\mathbf{C}^n$ 的一个无关组，把 Gram-Schmidt 过程(0.6.4)应用于 A 的各列，用矩阵记号描述所得结果正是 A 的 QR 分解．Gram-

Schmidt算法的自然推广使同样的矩阵记号描述能够应用于 A 的各列可能是相关的一般情形. 设 $A=[a_1\cdots a_m]$写成了以它的列 $a_i\in\mathbf{C}^n$ 为子块的分块形式. 如果 $a_1=0$, 令 $q_1=0$; 否则令 $q_1=a_1/(a_1^*a_1)^{1/2}$. 对每个 $k=2, 3, \cdots, m$, 如同在通常的 Gram-Schmidt 过程中那样, 算出 [112]

$$y_k=a_k-\sum_{i=1}^{k-1}(q_i^*a_k)q_i.$$

如果 $y_k=0$(它成立, 当且仅当 a_k 是 $a_1, a_2, \cdots, a_{k-1}$ 的线性组合), 则令 $q_k=0$; 否则令 $q_k=y_k/(y_k^*y_k)^{1/2}$. 这时, 向量 $q_1, \cdots, q_m$ 是正交组, 其中的每个元素或者是单位向量, 或者是零向量. 每个向量 q_j 是 $a_1, \cdots, a_j$ 的线性组合, 反过来, 上述做法保证每个列 a_j 是 $q_1, \cdots, q_j$ 的线性组合. 因此存在纯量 r_{kj} 使得

$$a_j=\sum_{k=1}^{j}r_{kj}q_k, \quad j=1,2,\cdots,m. \tag{2.6.2}$$

如果对所有 $k>j$, 令 $r_{kj}=0$, 而对于所有使 $q_i=0$ 的每个 i, 对所有 $j=1, 2, \cdots, m$, 令 $r_{ij}=0$, 那么, 经过上述过程, 上三角矩阵 $R=[r_{ij}]\in M_m$ 和向量 $q_1, q_2, \cdots, q_m$ 由 $a_1, a_2, \cdots, a_m$ 所确定. 矩阵 $Q\equiv[q_1, \cdots, q_m]\in M_{n,m}$ 有正交列(其中有些可能是零), 因而(2.6.2)表明 $A=QR$.

如果 rank $A=m$, 则 Q 有标准正交列, 因而所要求的分解形式已经得到. 特别地, 如果 $m=n$, 且 A 是非奇异矩阵, 那么, 根据(2.1.4e), Q 必须是酉矩阵, 而非奇异矩阵 $R=Q^*A$ 的诸对角元必须是非零的. 在这种情形, 因为要求 R 是上三角矩阵, 故 q_1 是 a_1 的纯量倍数, 并且对于 $i=2, \cdots, n$, q_i 位于这样一个一维空间内, 它在由 $a_1, \cdots, a_i$ 生成的空间中是 $a_1, \cdots, a_{i-1}$ 生成空间的正交补. 因此, 除了差一个绝对值为 1 的比例因子以外, 每个 q_i 是唯一确定的. 于是用 $R'\equiv\text{diag}(|r_{11}|/r_{11}, \cdots, |r_{mm}|/r_{mm})R$ 代替 R 以及用 $Q'\equiv Q\,\text{diag}(r_{11}/|r_{11}|, \cdots, r_{mm}/|r_{mm}|)$代替 Q, 便给出唯一的分解 $A=Q'R'$, 这正是定理的叙述中所断言的.

如果 A 的各列不是无关的, 取由 Q 的非零列组成的(标准正交)组, 然后把它扩充为 $\mathbf{C}^n$ 的一个标准正交基; 用上述方法得来的新向量记作 $z_1, z_2, \cdots, z_p$. 现在, 用 z_1 代替 Q 的第 1 个零列, 用 z_2 代替第 2 个零列, 如此做下去, 直到所有零列都被这种方式所取代; 用 Q'表示所得到的矩阵. 因为 Q'的新列与 R 的零行相对应, 于是 Q'有标准正交列, 且 $QR=Q'R$, 因此 $A=Q'R$ 是所要求的分解式.

如果 A 是实矩阵, 注意到所有必要的运算都可经实的运算来完成, 因而所得到的因子是实矩阵. □ [113]

练习 如果 $A\in M_{n,m}$, 且 $n\leqslant m$, 证明, A 可以分解成 $A=LP$, 其中, $L\in M_n$ 是下三角矩阵, 而 $P\in M_{n,m}$具有标准正交行; 至于其余的论断, 这个分解有与(2.6.1)相平行的叙述.

练习 证明, 任一个形如 $B=A^*A$ 的矩阵 $B\in M_n$(其中 $A\in M_n$)都可以写成 $B=LL^*$, 其中 $L\in M_n$ 是具有非负对角元的下三角矩阵. 并证明, 如果 A 是非奇异矩阵, 那么这个分解是唯一的. 称这个分解为 B 的 Cholesky 分解; 每个正定矩阵都可以按这种方式分解(见第 7 章). **提示**: 写出 $A=QR$.

QR 分解有重要的数值意义[例如, (2.6.3)], 然而作为一种理论方法, 它同样是有价值的. 例如, 已知矩阵 $A\in M_n$ 经酉相似的上三角化, 可以直接从它的通常的相似上三角化得到.

如果 $S^{-1}AS=T$ 是上三角矩阵，且如(2.6.1)中，$S=QR$，那么 $R^{-1}Q^*AQR=T$ 且 $Q^*AQ=RTR^{-1}$，而它作为上三角矩阵的乘积仍是上三角矩阵．用同样的方式可以推出，有关同时三角化的定理，例如(2.4.15)，实际上是同时酉三角化定理．这就是说，如果 M_n 中的某个矩阵族可经相似同时三角化，那么它同样可经酉等价同时三角化．

下面，不加证明地叙述计算特征值的 QR 算法，并简要地说明它的某些特性．

2.6.3 QR 算法　设 $A_0\in M_n$ 已知．写出 $A_0=Q_0R_0$，其中 Q_0 和 R_0 如(2.6.1)所示，然后定义 $A_1=R_0Q_0$．再写出 $A_1=Q_1R_1$，其中 Q_1 是酉矩阵，而 R_1 是上三角矩阵，这样继续做下去，一般地，如果因子 $A_k=Q_kR_k$，就定义 $A_{k+1}=R_kQ_k$．

练习　证明，用 QR 算法得到的每个 A_k 酉等价于 A_0，$k=1,2,\cdots$．

在某些情形(例如，如果 A_0 的所有特征值有不同的绝对值)，当 $k\to\infty$ 时，QR 迭代值 A_k 将收敛于一个上三角矩阵．因为这个上三角矩阵酉等价于 A_0，所以 A_0 的各特征值便被计算出来．

如果 A_0 是实矩阵，QR 迭代值 A_k 可以通过实的运算来计算．如果 A_0 有非实特征值，那么就不能指望 QR 迭代值会收敛于一个上三角矩阵，这是因为这个上三角极限必须是实矩阵．不过，在某些情形，可以选取迭代值 A_k，使它们收敛于一个具有 1×1 和 2×2 主对角子块的实分块上三角矩阵．而出现这种情形的充分条件是，除了各个成双成对的非实复共轭特征值有相同的模以外，A_0 的所有实特征值都有不同的模．因为分块三角矩阵的各特征值是诸对角子块特征值的集合之并，所以 A_0 的特征值是作为 QR 迭代值 A_k 的分块三角极限的 1×1 对角元以及该极限的 2×2 对角子块的特征值(的复共轭对)而出现的，它们可以很容易地用实数运算和二次公式计算出来．[114]

2.6.4 例　下述例子说明，QR 算法不一定总收敛于一个三角矩阵．设

$$A=\begin{bmatrix}0&1\\-1&0\end{bmatrix}.$$

于是，$\sigma(A)=\{\pm i\}$，特征值没有不同的模．如果 $A_0=A$，这个算法的一个可能的序列是

$$A_0=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}-1&0\\0&1\end{bmatrix},$$

$$A_1=\begin{bmatrix}-1&0\\0&1\end{bmatrix}\begin{bmatrix}0&1\\1&0\end{bmatrix}=\begin{bmatrix}0&-1\\1&0\end{bmatrix}=\begin{bmatrix}0&1\\-1&0\end{bmatrix}\begin{bmatrix}-1&0\\0&-1\end{bmatrix},$$

$$A_2=\begin{bmatrix}-1&0\\0&-1\end{bmatrix}\begin{bmatrix}0&1\\-1&0\end{bmatrix}=A_0.$$

另一个是

$$A_0=A_0\begin{bmatrix}1&0\\0&1\end{bmatrix}=A_1.$$

在这两种情形都出现了循环，因而序列$\{A_k\}$不收敛于上三角矩阵．但是，存在$\{A_k\}$的一个选择，它收敛于一个分块上三角矩阵．

习题

1. 设 x_1，…，x_m 是 $\mathbf{C}^n$ 中的给定无关向量组，且设 $X\equiv[x_1x_2\cdots x_m]\in M_{n,m}$. 假定对向量 x_1，…，x_m 实施(0.6.4)中所描述的 Gram-Schmidt 过程得到标准正交组 z_1，…，z_m 且设 $Z\equiv[z_1\cdots z_m]\in M_{n,m}$. (a)对 $k=1$，2，…，m，设 $Z_k=[z_1z_2\cdots z_kx_{k+1}x_{k+2}\cdots x_m]$，其中 z_k 是经 Gram-Schmidt 过程的第 k 步而得到的单位向量，因而 $Z_m=Z$. 证明，$Z_1=X\Delta_1$，$Z_2=Z_1\Delta_2$，…，$Z_m=Z_{m-1}\Delta_m$. 其中，每个 Δ_i 都是非奇异上三角矩阵，而 $\Delta_i=I+$一个除第 i 列以外其余各列全为零的上三角矩阵. (b)设 $T_k=\Delta_1\Delta_2\cdots\Delta_k$，$k=1$，2，…，$m$. 说明 T_k 是上三角矩阵且 $Z_k=XT_k$，$k=1$，2，…，m. 设 $T=T_m$，因而 $Z=XT$. (c)矩阵 T 与定理(2.6.1)证明中的上三角矩阵 R 之间的相互关系是什么？(d)证明，对 $j=k+1$，$k+2$，…，m，Z_j 和 T_j 的前 k 列不会变动，因此，Gram-Schmidt 过程的第 k 步就产生最终得到的矩阵 Z 和 T 的第 k 列.

115

2. 设 x_1，…，x_m 是 $\mathbf{C}^n$ 中给定的线性无关向量组，又设 $X\equiv[x_1\cdots x_m]\in M_{n,m}$. 考虑下面的算法：

Ⅰ. 令 $Z\equiv X$，且记 $Z=[z_1\cdots z_m]$，一开始先让每列 $z_i=x_i$，$i=1$，…，m.

Ⅱ. 对 $k=1$，2，…，m，实施以下代换：

(ⅰ) 首先，列 z_k 用 $z_k/\langle z_k, z_k\rangle^{1/2}$代替；

(ⅱ) 对 $j=k+1$，$k+2$，…，m，每列 z_j 用 $z_j-\langle z_j, z_k\rangle z_k$ 代替.

我们用 $\langle x, y\rangle=y^*x$ 表示 $\mathbf{C}^n$ 上的普通内积. (a)证明这个过程的最后结果是矩阵 Z 具有标准正交列，它与习题 1 中用 Gram-Schmidt 过程得到的矩阵 Z 是相同的. (b)若 Z_k 表示该算法在第 k 步中止时矩阵 Z 的存储信息，$k=1$，2，…，m，证明 $Z_1=X\Delta_1$，$Z_2=Z_1\Delta_2$，…，$Z_m=Z_{m-1}\Delta_m$；这里，每个 Δ_i 是一个非奇异上三角矩阵，其形式是 $\Delta_i=I+$一个除第 i 行以外其余各行全为零的上三角矩阵. (c)设 $T_k\equiv\Delta_1\Delta_2\cdots\Delta_k$，$k=1$，2，…，$m$. 注意到 T_k 是上三角矩阵，而 $Z_k=XT_k$，$k=1$，2，…，m. 证明每个 T_k 的前 k 列与习题 1 中矩阵 T_k 的前 k 列是相同的，即使 Δ_k 和 Z_k 可能各不相同. 设 $T=T_m$. (d)证明，对 $j=k+1$，$k+2$，…，m，Z_j 和 T_j 的前 k 列不会变动，因此该算法的第 k 步就产生最终得到的矩阵 Z 和 T 的第 k 列. 这个算法称为修改的 Gram-Schmidt 过程；它通过重新调整计算得到与通常的 Gram-Schmidt 过程相同的结果. 虽然修改的算法与通常的 Gram-Schmidt 算法在数学上是等价的，但是对于数值计算来说，修改的 Gram-Schmidt 过程更受欢迎，因为它需要的存储较少，在遇到 X 的某些列接近平行这样一些难题时，它有助于得到一个能算出的 Z，这个 Z 的诸列要比用通常的 Gram-Schmidt 算法得到的 Z 更接近于正交. 在遇到难题时可以方便地通过列旋转策略使它的性能进一步改进：在实施步骤Ⅱ(ⅰ)之前，首先选作 z_k 的是一个余下的列 z_j，$j\geqslant k$，其长度的平方 $z_j^*z_j$ 最大. 在数值计算中每一步(不要求反演)实际上得到 Δ_i^{-1}，然后累积这些因子的乘积可以计算 X 的 QR 分解中的三角因子.

116

3. 说明如何用一系列由 Householder 变换组成的乘积来实现 QR 分解. 证明上述过程必须用 $n-1$ 个 Householder 变换来完成，且 Q 是这些变换的乘积. 一般认为，这个方法在计算上优于(2.6.1)证明中所采用的 Gram-Schmidt 算法.

4. 如果应用于 $A\in M_n$ 的 QR 算法收敛于一个上三角矩阵，可以怎样计算 A_0 的特征向量

呢？**提示**：需要解一个右边也是零的(奇异)三角形方程组.

5. 设将 QR 算法应用于某个矩阵 $A\in M_n$，且设 $\{A_k\}$ 是 QR 迭代值的序列. 如果序列收敛，且 $\lim_{k\to\infty}A_k=B$. 用选择原理(2.1.8)仔细说明为什么 B 酉等价于 A. 这为什么是重要的?

进一步阅读 关于 Gram-Schmidt、修改的 Gram-Schmidt 以及几个其他的正交化过程的另外的参考资料和详细的叙述，可参看[GVI]的 pp. 146-169. 关于 QR 算法和证明方法，以及另外的参考资料可见由 D. Watkins 写的综述，"Understanding the QR Algorithm," *SIAM Rev*. 24(1982)；也可参看[Ste]，427-440.

117

第3章 标 准 形

3.0 导引

在什么情形下两个矩阵是相似的？我们知道，相似矩阵有相同的迹、行列式、特征多项式和特征值，然而例子

$$A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} \quad 和 \quad B=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix} \tag{3.0.1}$$

说明，两个矩阵的上述四个量可以分别相等，但它们却不相似. 假如存在某个非奇异矩阵 $S\in M_n$ 使得 $A=SBS^{-1}=S0S^{-1}=0$，那么，因为 $A\neq 0$ 而得出矛盾.

练习 计算(3.0.1)中的两个矩阵的迹，行列式，特征多项式和特征值. 证明 $A^2=0$.

因为两个看上去很不相同的矩阵仍然可以相似，所以，一条确定两个矩阵是否相似的途径是，设想有某个具有指定形式的"简单"矩阵的集合，然后看这两个已知矩阵是否可以通过相似化成这些"简单"形式中的一个. 如果它们能做到，那么它们必定是相似的(因为相似关系是传递的和对称的). 什么样的"简单"形式能符合这个要求呢？

每个复矩阵 A(酉)相似于一个上三角矩阵，它的对角元(A 的特征值)可以按任一给定的次序(2.3.1)排列，因此，如果两个矩阵相似于同一个上三角矩阵，那么它们就是相似的. 不过，两个具有相同主对角元和不同的非对角元的上三角矩阵仍然可以相似. 这样一来，如果已经把两个已知矩阵化成两个具有相同主对角元的不相等的三角矩阵，还不能由此得出这两个矩阵不相似的结论. 这里有相当大的灵活性；为了识别是否相似，在一个上三角矩阵中有 $n(n+1)/2$ 个非零元(更确切地说"未必是零"的元)需要考察，这个数目太大了. 关于这种三角矩阵，没有唯一的形式.

如果说上三角矩阵类与所要求的形式相距甚远，那么对角矩阵类又如何呢？如果两个已知矩阵中的每一个都各相似于一个对角矩阵，那么，它们彼此相似，当且仅当两个对角矩阵有相同的对角元，其中相重对角元按重数计算而无需考虑它们的顺序. 其理由是：可以用置相似矩阵 PDP^T 按任意规定的次序给出对角矩阵 D 的主对角元. 虽然这解决了讨论上三角矩阵时所涉及的唯一性问题，然而，现在还有一个存在性问题：不是每个复矩阵都相似于一个对角矩阵.

练习 证明(3.0.1)中的矩阵不能对角化. **提示**：如果 $A=SAS^{-1}$，那么 $A=B$.

如果对每一个矩阵来求一个尽可能接近对角矩阵的上三角形式，而且还可用相似变换得到它，那么所得结果是 Jordan 标准形，这正是下一节要讨论的内容.

我们已经考虑过两个矩阵 A，$B\in M_n$ 的相似性，在矩阵理论中还存在其他有意义的等价关系. 例如，A 是否可以经酉相似或者只应用初等行和列的变换变成 B；如果 A 和 B 都是实矩阵，是否可经一个实相似使 A 相似于 B；如果 A 和 B 是 Hermite 矩阵，是否存在一个非奇异矩阵 $S\in M_n$，使得 $A=SBS^*$；如果 A 和 B 是对称矩阵，是否存在一个非奇异矩阵 $S\in M_n$

使得 $A=SBS^*$.

在上述每一个例子中，在一个矩阵集合上有一个等价关系，而我们对每两个矩阵是否在同一等价类中感兴趣．解决这个问题的途径是，求一个具有规定形式的代表矩阵的“简单”集合，它们各取自每个等价类中的一个矩阵．我们试图把每个已给矩阵化成它们中的一个．如果这一方法果真凑效，那么每个等价类实际上必须包含规定形式的一个代表矩阵(这在相似下对于对角矩阵集合是不成立的)，而且最好在每个类中只有一个代表矩阵(或者设想有一个由等价类的 [120] 代表矩阵组成的较小的且容易描述的集合)(这在相似下对于上三角矩阵集合是不成立的)．这样一个由代表矩阵组成的集合常常称为标准形，在本章里，要考虑几个标准形．在后几章的部分段落还将引出其他标准形．

3.1 Jordan 标准形：一个证明

Jordan 标准形是一个“近乎对角”矩阵的集合，称为 Jordan 矩阵，它包括对角矩阵．Jordan 矩阵具有性质：每个(在相似下的)复方阵的等价类包含一个 Jordan 矩阵，并且按一种显而易见的方式使得在同一个等价类的任意两个 Jordan 矩阵本质上是相同的．一个与已知矩阵相似的 Jordan 矩阵称为该矩阵的 Jordan 标准形(或者有时也称为 Jordan 法式)．只要知道了一个矩阵的 Jordan 标准形，那么关于该矩阵(即线性变换)的所有线性代数的信息一看就清楚了．

3.1.1 定义 Jordan 块 $J_k(\lambda)$ 是指具有形状：

$$J_k(\lambda)=\begin{bmatrix}\lambda & 1 & & & 0\\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1\\ 0 & & & & \lambda\end{bmatrix} \tag{3.1.2}$$

的 $k\times k$ 上三角矩阵．在上对角线上有 $k-1$ 项“+1”；纯量 λ 在主对角线上出现 k 次．其他所有元素都是零，且 $J_1(\lambda)=[\lambda]$．一个 Jordan 矩阵 $J\in M_n$ 是诸 Jordan 块的一个直和

$$J=\begin{bmatrix}J_{n_1}(\lambda_1) & & & 0\\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & J_{n_k}(\lambda_k)\end{bmatrix},\quad n_1+n_2+\cdots+n_k=n, \tag{3.1.3}$$

其中，各阶数 n_i 可以相同，而值 λ_i 未必不同．

应指出的是，如果(3.1.3)中的每个 Jordan 块 $J_{n_i}(\lambda_i)$ 都是一维的，即，所有 $n_i=1$ 且 $k=n$，那么 Jordan 矩阵 J 是对角矩阵．如果(3.1.3)中的任一 Jordan 块 $J_m(\lambda)$ 有 $m>1$ 那么 J 不仅不是对角矩阵，它甚至不能对角化．假如 $J_m(\lambda)=SAS^{-1}$ 且 A 是对角矩阵，那么必须有 $A=$ [121] $\mathrm{diag}(\lambda,\lambda,\cdots,\lambda)=\lambda I$．于是 $J_m(\lambda)-\lambda I=SAS^{-1}-\lambda I=\lambda I-\lambda I=0$，因而如果 $m>1$，这种情

形不会出现．相应于每个单独的 Jordan 块，都有 J 的一个特征向量；它是属于 J 中每个 $J_m(\lambda)$ 的第一个对角元的标准基向量．

本节的主要结果是，每个复矩阵都相似于一个实质上是唯一的 Jordan 矩阵．我们将通过三步来得到最终的结论．

第一步　注意到每个复矩阵相似于一个上三角矩阵，它的诸特征值按一个规定的顺序出现在主对角线上；这正是 Schur 三角化定理(2.3.1)．

第二步　然后证明，一个上三角矩阵可经相似变换成一个分块对角矩阵，其中每个单独的对角子块的所有对角元都相等[像 Jordan 块(3.1.2)一样]．这是定理(2.4.8)．

第三步　最后证明，一个其主对角元都相等的上三角矩阵相似于若干 Jordan 块(3.1.2)的直和．

只要证明了最后一个结论，就可以通过复合每一步所必需的相似变换将任一复矩阵化成 Jordan 标准形．

另外，我们也注意到以下事实：如果一个矩阵是实的，且仅有实特征值，那么可以用实相似把它化成 Jordan 标准形．为此，我们想到了(2.3.1)，那是说，如果实矩阵 A 只有实特征值，那么存在一个实酉(实正交)矩阵 U，使得 U^TAU 是上三角矩阵，因而它只有实元素．此外(2.4.8)的证明说明，如果上三角矩阵 A 是实的，那么存在一个实相似矩阵 S，使得 $S^{-1}AS$ 是一个(实)分块对角矩阵，其中，每个对角子块是上三角矩阵，且都具有相等的主对角元．因此，余下只需要证明第三步是可以实现的；并且，如果从一个具有相同的实主对角元的实上三角矩阵开始实施变换，那么，把它化成一个 Jordan 块的直和的相似矩阵可以取实矩阵．

在证明第三步是可以实现的过程中，下述引理是有用的．它的证明其实就是直接计算．

3.1.4 引理　设 $k\geqslant 1$ 是已知的，且假定有 Jordan 块

$$J_k(0)=\begin{bmatrix}0 & 1 & & 0\\ & & \ddots & \\ & \ddots & & 1\\ 0 & & & 0\end{bmatrix}.$$

122

那么

$$J_k^T(0)J_k(0)=\begin{bmatrix}0 & 0\\ 0 & I_{k-1}\end{bmatrix},\quad \text{且如果 } p\geqslant k,\text{则 } J_k(0)^p=0.$$

另外，$J_k(0)e_{i+1}=e_i$，$i=1,2,\cdots,k-1$，且 $[I-J_k^T(0)J_k(0)]x=(x^Te_1)e_1$，这里，$I_{k-1}\in M_{k-1}$ 是单位矩阵，e_i 是第 i 个标准单位基向量，且 $x\in C^n$．

现在证明，第三步中的简化总是可以做到的．我们知道一个严格上三角矩阵是其主对角线上只有零元的三角矩阵．同时注意到一个具有相等主对角元的上三角矩阵是单位矩阵的一个纯量倍数加上一个严格上三角矩阵．

3.1.5 定理　设 $A\in M_n$ 是严格上三角矩阵．存在一个非奇异矩阵 $S\in M_n$ 和整数 $n_1,n_2,\cdots,n_m$，其中 $n_1\geqslant n_2\geqslant\cdots\geqslant n_m\geqslant 1$ 且 $n_1+n_2+\cdots+n_m=n$，使得

$$A = S\begin{bmatrix} J_{n_1}(0) & & & 0 \\ & J_{n_2}(0) & & \\ & & \ddots & \\ 0 & & & J_{n_m}(0) \end{bmatrix}S^{-1}. \tag{3.1.6}$$

如果 A 是实矩阵，则矩阵 S 也可以取实矩阵.

证明： 如果 $n=1$，$A=[0]$，结论是明显的. 对 n 作归纳法，且假定对阶数小于 n 的所有严格上三角矩阵，结论已经证明. 把 A 块分成

$$A = \begin{bmatrix} 0 & a^T \\ 0 & A_1 \end{bmatrix},$$

其中 $a\in \mathbf{C}^{n-1}$ 且 $A_1\in M_{n-1}$ 是严格上三角矩阵. 根据归纳假设，存在非奇异矩阵 $S_1\in M_{n-1}$，使得 $S_1^{-1}A_1S_1$ 有所要求的形式(3.1.6)：即

$$S_1^{-1}A_1S_1 = \begin{bmatrix} J_{k_1} & & & 0 \\ & J_{k_2} & & \\ & & \ddots & \\ 0 & & & J_{k_s} \end{bmatrix} = \begin{bmatrix} J_{k_1} & 0 \\ 0 & J \end{bmatrix}, \tag{3.1.7}$$

[123] 其中 $k_1\geqslant k_2\geqslant\cdots\geqslant k_s\geqslant 1$，$k_1+k_2+\cdots+k_s=n-1$，$J_{k_i}=J_{k_i}(0)$，且

$$J = \begin{bmatrix} J_{k_2} & & 0 \\ & \ddots & \\ 0 & & J_{k_s} \end{bmatrix} \in M_{n-k_1-1}.$$

注意到 J 中没有一个对角 Jordan 块的阶数超过 k_1，所以，根据引理(3.1.4)，$J^{k_1}=0$，经简单计算可知：

$$\begin{bmatrix} 1 & 0 \\ 0 & S_1^{-1} \end{bmatrix} A \begin{bmatrix} 1 & 0 \\ 0 & S_1 \end{bmatrix} = \begin{bmatrix} 0 & a^TS_1 \\ 0 & S_1^{-1}A_1S_1 \end{bmatrix}. \tag{3.1.8}$$

使分块 $a^TS_1=[a_1^T\ a_2^T]$与(3.1.7)最右边的分块相一致；即，$a_1\in\mathbf{C}^{k_1}$，且 $a_2\in\mathbf{C}^{n-1-k_1}$，然后把(3.1.8)写成

$$\begin{bmatrix} 1 & 0 \\ 0 & S_1^{-1} \end{bmatrix} A \begin{bmatrix} 1 & 0 \\ 0 & S_1 \end{bmatrix} = \begin{bmatrix} 0 & a_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix}.$$

现在考察这个矩阵的下述相似矩阵：

$$\begin{bmatrix} 1 & -a_1^TJ_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}\begin{bmatrix} 0 & a_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix}\begin{bmatrix} 1 & a_1^TJ_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} 0 & a_1^T(I-J_{k_1}^TJ_{k_1}) & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix} = \begin{bmatrix} 0 & (a_1^Te_1)e_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix}, \tag{3.1.9}$$

其中，用到了引理(3.1.4)中的恒等式$(I-J_k^TJ_k)x=(x^Te_1)e_1$. 现在有两种可能性，这取决于$a_1^Te_1=0$还是$a_1^Te_1\neq 0$.

如果$a_1^Te_1\neq 0$，那么

$$\begin{bmatrix}1/a_1^Te_1 & 0 & 0\\ 0 & I & 0\\ 0 & 0 & (1/a_1^Te_1)I\end{bmatrix}\begin{bmatrix}0 & (a_1^Te_1)e_1^T & a_2^T\\ 0 & J_{k_1} & 0\\ 0 & 0 & J\end{bmatrix}\begin{bmatrix}a_1^Te & 0 & 0\\ 0 & I & 0\\ 0 & 0 & a_1^Te_1I\end{bmatrix}$$

$$=\begin{bmatrix}0 & e_1^T & a_2^T\\ 0 & J_{k_1} & 0\\ 0 & 0 & J\end{bmatrix}\equiv\begin{bmatrix}\tilde{J} & e_1a_2^T\\ 0 & J\end{bmatrix}.$$

注意

$$\tilde{J}=\begin{bmatrix}0 & e_1^T\\ 0 & J_{k_1}\end{bmatrix}=J_{k_1+1}(0)$$

是一个具有零主对角线的k_1+1阶 Jordan 块. 利用性质$\tilde{J}e_{i+1}=e_i$，$i=1,2,\cdots,k_1$，容易证明 [124]

$$\begin{bmatrix}I & e_2a_2^T\\ 0 & I\end{bmatrix}\begin{bmatrix}\tilde{J} & e_1a_2^T\\ 0 & J\end{bmatrix}\begin{bmatrix}I & -e_2a_2^T\\ 0 & I\end{bmatrix}=\begin{bmatrix}\tilde{J} & -\tilde{J}e_2a_2^T+e_1a_2^T+e_2a_2^TJ\\ 0 & J\end{bmatrix}$$

$$=\begin{bmatrix}\tilde{J} & e_2a_2^TJ\\ 0 & J\end{bmatrix},$$

然后可以递归地计算一系列相似的矩阵

$$\begin{bmatrix}I & e_{i+1}a_2^TJ^{i-1}\\ 0 & I\end{bmatrix}\begin{bmatrix}\tilde{J} & e_ia_2^TJ^{i-1}\\ 0 & J\end{bmatrix}\begin{bmatrix}I & -e_{i+1}a_2^TJ^{i-1}\\ 0 & I\end{bmatrix}=\begin{bmatrix}\tilde{J} & e_{i+1}a_2^TJ^i\\ 0 & J\end{bmatrix}$$

$$i=2,3,\cdots.$$

因为$J^{k_1}=0$，我们看出，在这一系列相似的矩阵中最多经k_1步便可使非对角子块变为零. 因此得出，A相似于矩阵

$$\begin{bmatrix}\tilde{J} & 0\\ 0 & J\end{bmatrix},$$

它正是要求的严格上三角 Jordan 矩阵形式.

如果$a_1^Te_1=0$，那么(3.1.9)说明，A相似于矩阵

$$\begin{bmatrix}0 & 0 & a_2^T\\ 0 & J_{k_1} & 0\\ 0 & 0 & J\end{bmatrix},$$

而它又置换相似于矩阵

$$\begin{bmatrix}J_{k_1} & 0 & 0\\ 0 & 0 & a_2^T\\ 0 & 0 & J\end{bmatrix}. \qquad (3.1.10)$$

根据归纳假设，存在非奇异矩阵$S_2\in M_{n-k_1}$，使得：

$$S_2^{-1}\begin{bmatrix}0 & a_2^T\\ 0 & J\end{bmatrix}S_2 = \tilde{J} \in M_{n-k}$$

是一个具有零主对角线的 Jordan 矩阵. 因此，矩阵(3.1.10)，从而 A 本身相似于

$$\begin{bmatrix}J_{k_1} & 0\\ 0 & \tilde{J}\end{bmatrix},$$

它就是所要求的 Jordan 标准形式，只是诸对角 Jordan 块可能不按非增的顺序排列. 若有必要，经一个分块置换相似便可得到所要求的形式.

[125] 最后，我们看到，如果 A 是实矩阵，那么在这个证明中所采用的所有相似矩阵都可选为实的，于是经过一个实相似，A 便相似于所要求的 Jordan 矩阵. □

为建立起 Jordan 标准形，定理(3.1.5)实质上完成了所约定的程序的第三步. 我们注意到，如果

$$A = \begin{bmatrix}\lambda & & & *\\ & \lambda & & \\ & & \ddots & \\ 0 & & & \lambda\end{bmatrix}$$

是所有对角元都等于 λ 的上三角矩阵，那么，$A_0 = A - \lambda I$ 是严格上三角矩阵. 如果 $S\in M_n$ 是非奇异矩阵，且 $S^{-1}A_0S$ 是由(3.1.5)所确保的若干个基本 Jordan 块 $J_{n_i}(0)$ 的一个直和，则 $S^{-1}AS = S^{-1}A_0S+\lambda I$ 是若干个基本 Jordan 块 $J_{n_i}(\lambda)$ 的直和. 在(2.3)节中实施的第一步和第二步连同第三步恰好证明了下述 Jordan 标准形定理的存在性部分：

3.1.11 定理 设 $A\in M_n$ 是已知的复矩阵，则存在非奇异矩阵 $S\in M_n$，使得

$$A = S\begin{bmatrix}J_{n_1}(\lambda_1) & & & 0\\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & J_{n_k}(\lambda_k)\end{bmatrix}S^{-1} = SJS^{-1}, \qquad (3.1.12)$$

且 $n_1+n_2+\cdots+n_k=n$. 如果不计各对角 Jordan 块的排列顺序，A 的 Jordan 矩阵是唯一的. 特征值 λ_i，$i=1, 2, \cdots, k$，不一定不相同. 如果 A 是一个仅有实特征值的实矩阵，那么相似矩阵 S 可以取实矩阵.

证明：除了唯一性外，其余的论断都已证明. 如果 A，$B\in M_n$ 相似，则对于任一纯量 $\lambda\in\mathbf{C}$ 和任意指数 $m=1, 2, \cdots$，矩阵 $(A-\lambda I)^m$ 和 $(B-\lambda I)^m$ 也相似；特别地，它们的秩相等. 因此，只需证明，位于 Jordan 矩阵 $J\in M_n$ 对角线上的一组 Jordan 块(包括相重的子块)可由有限多个整数 $\operatorname{rank}(J-\lambda I)^m$ 完全确定，其中 $m=1, 2, \cdots, n$，$\lambda\in\sigma(J)$.

首先考虑形如(3.1.2)的一个 Jordan 块 $J_k(\mu)\in M_k$ 的情形，其中 $\mu\in\mathbf{C}$ 是给定的且 $m\geq 1$. 如果 $\mu\neq 0$，则 $\operatorname{rank} J_k(\mu)^m = \operatorname{rank} J_k(\mu)^{m+1} = k$，因而 $\operatorname{rank} J_k(\mu)^m - \operatorname{rank} J_k(\mu)^{m+1} = 0$. 如果 [126] $\mu=0$ 且 $m\geq k$，则 $J_k(0)^m = J_k(0)^{m+1}=0$，因而也有 $\operatorname{rank} J_k(0)^m - \operatorname{rank} J_k(0)^{m+1}=0$. 最后，如果 $\mu=0$ 且 $m<k$，则 $\operatorname{rank} J_k(0)^m - \operatorname{rank} J_k(0)^{m+1}=1$.

其次，设 $J\in M_n$ 是形如(3.1.3)的 Jordan 矩阵，又设 $\lambda\in\sigma(J)$，并且对于 $m=1, 2, \cdots$，定义 $r_m(\lambda)\equiv\operatorname{rank}(J-\lambda I)^m$；且令 $r_0(\lambda)\equiv n$. 从前述关于一个子块情形的分析可知，差 $d_m(\lambda)\equiv r_{m-1}(\lambda)-r_m(\lambda)$ 等于 J 中所有阶数 $k\geqslant m$ 的诸 Jordan 块 $J_k(\lambda)$ 的总数，且对所有 $m>n$ 有 $d_m(\lambda)=0$. 因此，J 中恰好是阶数 $k=m$ 的诸 Jordan 块的个数等于 $d_m(\lambda)-d_{m+1}(\lambda)=r_{m-1}(\lambda)-2r_m(\lambda)+r_{m+1}(\lambda)$，其中 $m=1, 2, \cdots, n$. □

练习 设 $A\in M_n$ 有 Jordan 标准形 J. 设 λ 是 A 的一个特征值，其代数重数为 ν，又设 b_k 表示 J 中阶数为 k 的诸 Jordan 块 $J_k(\lambda)$ 的个数，其中 $k=1, \cdots, n$. 如果当 $m\geqslant 1$ 时 $r_m(\lambda)=\operatorname{rank}(A-\lambda I)^m$，而 $r_0(\lambda)\equiv n$，证明：

(a) $r_m(\lambda)$ 和 b_i 满足线性方程组

$$r_m(\lambda)=n-\nu+\sum_{i=m+1}^{n}(i-m)b_i,\quad m=0,1,\cdots,n-1.$$

(b) 该方程组有唯一解. (c) 其解为 $b_m=r_{m-1}(\lambda)-2r_m(\lambda)+r_{m+1}(\lambda)$，$m=1, 2, \cdots, n$. 其中 $r_{n-1}(\lambda)=r_n(\lambda)=n-\nu$.

为了有 Jordan 标准形的一个标准表示(3.1.2)，我们约定选取 A 的各不相同的特征值 λ_1，λ_2，$\cdots$，λ_k 的某个顺序，并且首先给出相应于 λ_1 的所有 Jordan 块；然后是相应于 λ_2 的那些 Jordan 块，如此等等. 相应于每个不同特征值的诸 Jordan 块，按递减(非增)的顺序给出，首先是最大的子块，随后是仅次于最大的子块，等等. 因为，相应于同一特征值的多重同阶子块是完全相同的，所以，一旦给定了诸特征值的顺序，这个表示就给出了个唯一确定的 Jordan 标准形. M_n 中的矩阵的每个相似等价类包含一个且只包含一个这样的 Jordan 标准形.

虽然推导 Jordan 标准形的过程是一个明确的算法，它原则上可以用来计算一个已知矩阵的 Jordan 标准形，但是绝对不能指望用计算机对它自动实施数值计算. 令人遗憾的事实是，使用计算机不仅可能得出虚假的结果，而且实际上并没有一个计算 Jordan 标准形的稳定的数值方法. 这只要举一个简单的例子就清楚了.

如果 $A_\varepsilon=\begin{bmatrix}\varepsilon&0\\1&0\end{bmatrix}$，且 $\varepsilon\neq 0$. 那么 $A_\varepsilon=S_\varepsilon J_\varepsilon S_\varepsilon^{-1}$，其中 $S_\varepsilon=\begin{bmatrix}0&\varepsilon\\1&1\end{bmatrix}$，且 $J_\varepsilon=\begin{bmatrix}0&0\\0&\varepsilon\end{bmatrix}$. 如果令 $\varepsilon\to 0$，那么，$J_\varepsilon\to\begin{bmatrix}0&0\\0&0\end{bmatrix}$，而它不可能是非零矩阵 $A_0=\begin{bmatrix}0&0\\1&0\end{bmatrix}$的 Jordan 标准形. 事实上，$A_0$ 有 $\begin{bmatrix}0&1\\0&0\end{bmatrix}$作为它的 Jordan 标准形. 因为一个矩阵的 Jordan 形未必是该矩阵的各元素的一个连续函数，可能有这种情形，一个矩阵的各元的一个小的变化会引起 Jordan 标准形的各元的一个大的变化. 不能指望用稳定的方法计算这样的对象. 因此在数值应用中，几乎没有用到 Jordan 标准形. [127]

尽管有这样的局限性，Jordan 标准形还是值得认真了解的，它为透彻理解矩阵提供了丰富的源泉. 作为一般的技巧，当我们要证明矩阵的有关结论时，不妨先考虑能否对对角矩阵证明这一结论. 如果这是可行的，那么就(利用任一复矩阵可以用一个可对角化的矩阵任意接近的事实)看一看某个极限论证是否可以一般地证明该结论. 如果这不见效，或者想避开分析论证，那么下一步便可设法对上三角矩阵或 Jordan 矩阵来证明这一结论.

每个矩阵都相似于形如(3.1.12)那样的矩阵，其中，诸 Jordan 块中的所有项"+1"都用

$\varepsilon\neq0$代替，ε可以取任意小的值．了解这一点有时是有用的．

3.1.13 推论 设$A\in M_n$是给定的复矩阵，又设$\varepsilon>0$是已知的．那么存在一个非奇异矩阵$S=S(\varepsilon)\in M_n$，使得

$$A=S\begin{bmatrix}J_{n_1}(\lambda_1,\varepsilon) & & & 0\\ & J_{n_2}(\lambda_2,\varepsilon) & & \\ 0 & & \ddots & \\ & & & J_{n_k}(\lambda_k,\varepsilon)\end{bmatrix}S^{-1}, \qquad (3.1.14)$$

$$n_1+n_2+\cdots+n_k=n$$

且

$$J_m(\lambda,\varepsilon)\equiv\begin{bmatrix}\lambda & \varepsilon & & & 0\\ & \lambda & \varepsilon & & \\ & & \ddots & \ddots & \\ & & & \ddots & \varepsilon\\ 0 & & & & \lambda\end{bmatrix}\in M_m.$$

如果A是只具有实特征值的实矩阵，那么S可以取实矩阵．

证明：首先求非奇异矩阵$S_1\in M_n$，使得$S_1^{-1}AS_1$是Jordan标准形(如果A是实的，且只有实特征值，那么取实S_1)．然后取$D_\varepsilon=\text{diag}(1,\ \varepsilon,\ \varepsilon^2,\ \cdots,\ \varepsilon^{n-1})$，再计算$D_\varepsilon^{-1}(S_1^{-1}AS_1)D_\varepsilon$．这个矩阵具有形状(3.1.14)，于是$S=S(\varepsilon)=S_1D_\varepsilon$符合定理的要求．□

128

习题

1．用详细的计算来证明引理(3.1.4)．

2．试用(3.1.11)的证明中的三个步骤，求

$$\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} \quad 和 \quad \begin{bmatrix}3 & 1 & 2\\ 0 & 3 & 0\\ 0 & 0 & 3\end{bmatrix}$$

Jordan标准形．

3．设$A\in M_n$是复矩阵，但只有实特征值．证明A相似于一个实矩阵，相似矩阵可以选取实矩阵吗？

进一步阅读 定理(3.1.11)的证明思想取自R. Fletcher and D. Sorensen, "An Algorithmic Derivation of the Jordan Canonical Form," *Amer. Math. Monthly* 90(1983), 12-16，文中另有参考文献．[Ste]从数值计算的观点讨论了Jordan标准形，并且给出了例子说明矩阵的元素产生扰动时其Jordan标准形的灵敏度．[Str]提供了一个好的证明．

3.2 Jordan标准形：若干论断和应用

3.2.1 Jordan矩阵的结构 Jordan矩阵

$$J=\begin{bmatrix}J_{n_1}(\lambda_1) & & & 0\\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & J_{n_k}(\lambda_k)\end{bmatrix},\quad n_1+n_2+\cdots+n_k=n \qquad (3.2.1.1)$$

有一个确定的结构，它使得该矩阵以及与其相似的任一矩阵的某些基本性质变得更加明显.

1. Jordan 块的个数 k(相同的子块计重复出现的次数)是 J 的线性无关特征值向量的个数.

2. 矩阵 J 可对角化，当且仅当 $k=n$.

3. 相应于一个已知特征值的 Jordan 块的个数是该特征值的几何重数. 它是相应的特征空间的维数. 相应于一个已知特征值的所有 Jordan 块的阶数之和是该特征值的代数重数. 129

4. 知道了诸特征值以及它们的代数重数和几何重数，一般不能完全确定一个 Jordan 矩阵. 我们还需知道相应于每个特征值的各 Jordan 块的阶数，相应于一个特征值 λ 的最大 Jordan 块的阶数是 λ 作为其极小多项式的一个根的重数(3.3.6).

5. 相应于某个特征值的诸 Jordan 块的阶数可以通过计算某些幂的秩来确定. 例如，如果

$$J=\left[\begin{array}{ccc|cc|cc|c} 2&1&0&&&&&\\ 0&2&1&&&0&&\\ 0&0&2&&&&&\\ \hline &&&2&1&&&\\ &&&0&2&&&\\ \hline &&&&&2&1&\\ &0&&&&0&2&\\ \hline &&&&&&&2 \end{array}\right].$$

经计算

$$J-2I=\left[\begin{array}{ccc|cc|cc|c} 0&1&0&&&&&\\ 0&0&1&&&0&&\\ 0&0&0&&&&&\\ \hline &&&0&1&&&\\ &&&0&0&&&\\ \hline &&&&&0&1&\\ &0&&&&0&0&\\ \hline &&&&&&&0 \end{array}\right],$$

$$(J-2I)^2=\left[\begin{array}{ccc|cc|cc|c} 0&0&1&&&&&\\ 0&0&0&&&0&&\\ 0&0&0&&&&&\\ \hline &&&0&0&&&\\ &&&0&0&&&\\ \hline &&&&&0&0&\\ &0&&&&0&0&\\ \hline &&&&&&&0 \end{array}\right]$$

130

和 $(J-2I)^3=0$. 因此，我们知道

$$(J-2I)^3=0,$$
$$\mathrm{rank}(J-2I)^2=1,$$
$$\mathrm{rank}(J-2I)=4,$$
$$(J-2I)\text{ 是 } 8\times 8 \text{ 矩阵}.$$

这组数足以确定 J 的分块结构. 事实上，$(J-2I)^3=0$ 说明最大子块的阶是 3. $(J-2I)^2$ 的秩就是 3 阶子块的个数，因而只有一个 3 阶子块. $(J-2I)$的秩是 3 阶子块的个数的 2 倍加上 2 阶子块的个数，因而有两个 2 阶子块. 1 阶子块的个数是 $8-(2\times 2)-3=1$. 可以把同样的手续应用于具有任意阶数的若干 Jordan 块的直和，只要它们是相应于同一个特征值的所有子块. 如果 J 是相应于特征值 λ 的这样一个直和，那么使$(J-\lambda I)^{k_1}=0$ 的最小整数 k_1 便是最大子块的阶数. $(J-\lambda I)^{k_1-1}$的秩是 k_1 阶子块的个数，$(J-\lambda I)^{k_1-2}$的秩是 k_1 阶子块的个数的 2 倍再加上 k_1-1 阶子块的个数，如此等等. 对于 $i=0, 1, 2, \cdots, k_1-1$，$(J-\lambda I)^{k_1-i}$的秩序列递归地确定出 J 中所有子块的阶数.

6. 在一般的 Jordan 矩阵(3.1.1.1)中的所有 Jordan 子块的阶数可以通过分析某些幂的秩来确定. 如果 λ_1 是 Jordan 矩阵 $J\in M_n$ 的特征值，那么，当我们取幂$(J-\lambda_1 I)$，$(J-\lambda_1 I)^2$，…时，只有相应于 λ_1 的诸 Jordan 块会变成零，因为 $J-\lambda_1 I$ 中的其他子块都有非零对角元. 最终，$(J-\lambda_1 I)^k$ 的秩将停止下降(不必考察任一 $k>n$)；使$(J-\lambda_1 I)^k$ 的秩达到它的极小值的那个最小的 k 值是相应于 λ_1 的最大子块的阶. 这个极小值称为特征值 λ_1 的指标. 对$(J-\lambda_1 I)$的各次幂组成的序列的秩进行分析，就足以确定相应于 λ_1 的诸 Jordan 块的阶数与个数，然后再对 λ_2，λ_3，等等作同样的分析，因而确定出所有的子块.

虽然上述各个论断是针对 Jordan 矩阵 J 作出的，然而，每一个论断也同样适用于任一相似于 J 的矩阵. 因此，如果已知矩阵 $A\in M_n$，那么 A 的 Jordan 标准形(而不是把 A 变换成 Jordan 标准形的相似)可以通过下述步骤来确定：

131

1. 求出 A 的所有不同的特征值，或许这可以通过求特征多项式的根来办到.

2. 对 A 的每个不同的特征值 λ_i，作$(A-\lambda_i I)^k$，其中 $k=1, 2, \cdots, n$，然后分析这些矩阵的秩所组成的序列便可看出 A 的相应于特征值 λ_i 的所有 Jordan 块的阶数与个数.

这个算法对于手算一些简单形式的小矩阵可能是有用的，但是用计算机计算就可能失效，因为确定一个矩阵的秩的过程本来就是一个不稳定的过程. 例如，$A_\epsilon=\begin{bmatrix}1&0\\0&\epsilon\end{bmatrix}$就可以使我们看清这一点：对所有 $\epsilon\neq 0$，A_ϵ 有秩 2，但对 $\epsilon=0$，它有秩 1.

例如，这个算法在下述情形是有用的. 假定有这样一个问题，要确定 Jordan 块 $J_k(0)\in M_k$ 的平方的 Jordan 标准形：

$$A=\begin{bmatrix}0&1&&0\\&&\ddots&\\&&\ddots&1\\&0&&0\end{bmatrix}^2=\begin{bmatrix}0&0&1&\cdots&0\\&&&\ddots&\vdots\\&&\ddots&&1\\&&\ddots&\ddots&\\&0&&\ddots&0\\&&&&0\end{bmatrix}.$$

A 的各特征值都是零，对 $m=\lceil(k+1)/2\rceil=(k+1)/2$ 的整数部分，$A^m=0$，而如果 $p=1, 2, \cdots, m-1$，则$A^p\neq 0$. 对于 $p=1, 2, \cdots, m-2$，每个幂 A^{p+1} 的秩比它的前一个幂 A^p 的秩少 2. 如果 k 是偶数，则 A^{m-1} 有秩 2，而如果 k 为奇数，则 A^{m-1} 有秩 1. 因此，$A=J_k^2(0)$ 的 Jordan 标准形是

$$\begin{bmatrix} J_m(0) & 0 \\ 0 & J_m(0) \end{bmatrix}, \quad k=2m \text{ 是偶数；}$$

以及

$$\begin{bmatrix} J_m(0) & 0 \\ 0 & J_{m-1}(0) \end{bmatrix}, \quad k=2m-1 \text{ 是奇数}.$$

这一结果对于确定一个已知矩阵是否有平方根是有用的. 比如，它说明不存在使 $A^2=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$的矩阵 $A\in M_2$.

3.2.2 常系数线性微分方程组 在理论上具有重要意义的 Jordan 标准形的一个应用是分析常系数微分方程组的解. 设 $A\in M_n$ 是已知的，且考虑一阶初值问题 [132]

$$\begin{aligned} x'(t) &= Ax(t), \\ x(0) &= x_0 \text{ 是已知的}, \end{aligned} \tag{3.2.2.1}$$

其中 $x(t)=[x_1(t), x_2(t), \cdots, x_n(t)]^T$，且撇号(′)表示对 t 微分. 如果 A 不是对角矩阵，则这个方程组是耦合的；即 $x_i(t)$不仅与 $x_i(t)$有关，而且还与向量 $x(t)$的其他分量有关. 这种耦合情形使问题变得难以解决，但是，如果可以把 A 变成对角(或近乎对角)的形式，那么耦合的影响可以减弱，或者甚至可以消除，因而可以很容易地求解. 如果 $A=SJS^{-1}$，其中 J 是 A 的 Jordan 标准形，那么(3.2.2.1)变成

$$\begin{aligned} y'(t) &= Jy(t), \\ y(0) &= y_0 \text{ 是已知的}, \end{aligned} \tag{3.2.2.2}$$

其中 $x(t)=Sy(t)$，且 $y_0=S^{-1}x_0$. 如果可以解出问题(3.2.2.2)，那么(3.2.2.1)的解的每个分量就是(3.2.2.2)的解的诸分量的一个线性组合，且这个线性组合是由 S 给出的.

如果 A 是可对角化的，那么 J 是对角矩阵，且(3.2.2.2)正好是一个形如 $y_k'(t)=\lambda_k y_k(t)$ 的非耦合方程组，它有解 $y_k(t)=y_k(0)e^{\lambda_k t}$. 如果特征值 λ_k 是实的，这是一个简单的幂，而如果 $\lambda_k=a_k+ib_k$ 是复数，它则是一个振动项 $y_k(t)=y_k(0)e^{a_k t}[\cos(b_k t)+i\sin(b_k t)]$.

如果 J 是非对角的，解就更复杂了. 相应于 J 中的不同 Jordan 块的 $y(t)$的诸分量是非耦合的. 因此只要考虑 J 是单独的 Jordan 块

$$J=\begin{bmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix} \in M_m$$

的情形就可以了. 方程组(3.2.2.2)是

$$y'_1(t)=\lambda y_1(t)\quad+y_2(t),$$
$$\vdots\qquad\vdots\qquad\vdots$$
$$y'_{m-1}(t)=\lambda y_{m-1}(t)+y_m(t),$$
[133]
$$y'_m(t)=\lambda y_m(t),$$

它可以由下往上直接求解. 从最后一个方程开始，得到

$$y_m(t)=y_m(0)e^{\lambda t},$$

于是

$$y'_{m-1}(t)=\lambda y_{m-1}(t)+y_m(0)e^{\lambda t}.$$

它有解

$$y_{m-1}(t)=y_m(0)te^{\lambda t}+y_{m-1}(0)e^{\lambda t},$$

于是可以把它用于下一个方程，它变成

$$y'_{m-2}(t)=\lambda y_{m-2}(t)+y_m(0)te^{\lambda t}+y_{m-1}(0)e^{\lambda t}.$$

它有解

$$y_{m-2}(t)=y_m(0)\frac{t^2}{2}e^{\lambda t}+y_{m-1}(0)te^{\lambda t}+y_{m-2}(0)e^{\lambda t},$$

如此等等. 显然解的每一个分量具有形式

$$y_k(t)=e^{\lambda t}q_k(t),$$

其中 $q_k(t)$是次数至多为 $m-1$ 的多项式.

从上述分析可以知道，对于任一给定的初始条件 x_0，问题(3.2.2.1)的解 $x(t)$的各个分量有形式

$$x_i(t)=p_1(t)e^{\lambda_1 t}+p_2(t)e^{\lambda_2 t}+\cdots+p_k(t)e^{\lambda_k t},$$

其中 $\lambda_1,\lambda_2,\cdots,\lambda_k$ 是 A 的不同的特征值，而每个 $p_i(t)$是多项式，它的次数严格小于相应于 λ_i 的最大 Jordan 块的阶数. 实特征值产生纯指数项，而复特征值可以产生混合指数和振动项.

3.2.3 矩阵与它的转置的相似性 每个 Jordan 块置换相似于它的转置，因为从相似性可以看出

$$\begin{bmatrix}0 & & 1\\ & \iddots & \\ 1 & & 0\end{bmatrix}\begin{bmatrix}\lambda & 1 & & 0\\ & \ddots & \ddots & \\ & & \ddots & 1\\ 0 & & & \lambda\end{bmatrix}\begin{bmatrix}0 & & 1\\ & \iddots & \\ 1 & & 0\end{bmatrix}=\begin{bmatrix}\lambda & & & 0\\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda\end{bmatrix}.$$

[134]

因此，如果 $A\in M_n$ 是已知的，且 $A=SJS^{-1}$是它的 Jordan 标准形，那么，A 相似于 J，而 J 相似于 J^T，J^T 又相似于 $A^T=(S^T)^{-1}J^T(S^T)$. 结论是，每个复矩阵相似于它的转置. 由此推出，一个复矩阵的行秩(线性无关行的最大个数)等于它的列秩(线性无关列的最大个数)，这是因为秩是相似不变量. 同时也推出，A 和 A^T 有相同的特征值. 不过，直接作出所有这些推论也是比较容易的.

对任意的域 **F**，情形也是这样，$M_n(\mathbf{F})$中的每个矩阵都可经 $M_n(\mathbf{F})$中的一个矩阵相似于它的转置；且不必假定 $\mathbf{F}=\mathbf{C}$. 实际上，相似矩阵可以取为对称矩阵.

3.2.4 交换矩阵和非减次矩阵 如果 $p(t)$是多项式，而 $A\in M_n$ 是已知矩阵，显然 $p(A)$与 A 可交换，这是一个有用的事实. 可是反过来又如何呢？如果 A，$B\in M_n$ 是已知的，且 A 与 B 可交换，B 必定是关于 A 的多项式吗？当然不是这样，因为如果取 $A=I$，则 A 与每个矩阵都可交换，$p(I)=p(1)I$ 不可能产生任何非纯量矩阵. 问题是 A 这一形式使它能与许多矩阵可交换，但又只允许它产生少数形如 $p(A)$的矩阵. 为了得到这方面的任何结果，必须寻求协调这两种能力之间的关系的某个折中办法.

3.2.4.1 定义 矩阵 $A\in M_n$ 叫做非减次矩阵，是指 A 的每个特征值有几何重数 1.

因为 Jordan 矩阵的一个特征值的几何重数等于相应于该特征值的 Jordan 块的个数，所以，一个矩阵是非减次的，当且仅当相应于每个不同的特征值恰好有一个 Jordan 块. 例如，如果 $A\in M_n$ 有 n 个不同的特征值，或者 A 只有一个特征值，而该特征值有几何重数 1，那么它是非减次的. 纯量矩阵不是非减次矩阵.

3.2.4.2 定理 设 $A\in M_n$ 是给定的非减次矩阵. 矩阵 $B\in M_n$ 与 A 可交换，当且仅当存在一个次数至多为 $n-1$ 的多项式 $p(\cdot)$使得 $B=p(A)$.

证明： 如果 $B=p(A)$，那么 B 一定与 B 可交换. 关于逆命题，设 $A=SJS^{-1}$是 A 的 Jordan 标准形. 如果 $BA=AB$，那么 $BSJS^{-1}=SJS^{-1}B$ 且$(S^{-1}BS)J=J(S^{-1}BS)$. 如果能够证明 $S^{-1}BS=p(J)$，那么 $B=Sp(J)S^{-1}=p(SJS^{-1})=p(A)$. 因此，只要假定 A 本身就是一个 Jordan 矩阵就可以了. 因为 A 是非减次的，所以可以假定

135

$$A=\begin{bmatrix} J_{n_1}(\lambda_1) & & & 0 \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & J_{n_k}(\lambda_k) \end{bmatrix},$$

其中 λ_1，λ_2，…，λ_k 是 A 的 k 个不同的特征值，如果把 B 写成与 A 的上述形式相同的分块形式 $B=(B_{ij})$，那么 $AB-BA$ 的相应非对角子块具有形式

$$J_{n_i}(\lambda_i)B_{ij}-B_{ij}J_{n_j}(\lambda_j)=0,\quad i\neq j.$$

因为特征值λ_i 和λ_j 是不同的，由此可以得出[见(2.4)节习题 9]，$B_{ij}=0$ 是这些方程的唯一解，因此矩阵 B 必须是分块对角矩阵

$$B=\begin{bmatrix} B_1 & & & 0 \\ & B_2 & & \\ & & \ddots & \\ 0 & & & B_k \end{bmatrix},$$

其中每个 $B_i\in M_{n_i}$. 交换性假设是说，对于所有 $i=1$，2，…，k，有 $B_iJ_{n_i}(\lambda_i)=J_{n_i}(\lambda_i)B_i$. 如果记 $J_{n_i}(\lambda_i)=\lambda_iI+N_i$，其中

$$N_i=\begin{bmatrix} 0 & 1 & & 0 \\ & & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix}\in M_{n_i},$$

那么这些恒等式变成 $B_iN_i=N_iB_i$，$i=1, 2, \cdots, k$. 直接计算可知，由于 N_i 的特殊形式每个 B_i 必须是具有 Toeplitz 形式(0.9.7)的上三角矩阵，即

$$B_i=\begin{bmatrix} b_1^{(i)} & b_2^{(i)} & \cdots & b_{n_i}^{(i)} \\ & b_1^{(i)} & \ddots & \vdots \\ 0 & & \ddots & b_2^{(i)} \\ & & & b_1^{(i)} \end{bmatrix}, \tag{3.2.4.3}$$

其中沿着各对角线的诸元素取常值. 如果可以构造出一些次数至多是 $n-1$ 的多项式 $p_i(t)$，且具有性质：对所有 $i\neq j$，$p_i(J_{n_j}(\lambda_j))=0$ 而 $p_i(J_{n_i}(\lambda_i))=B_i$，那么

136

$$p(t)=p_1(t)+\cdots+p_k(t)$$

将满足定理的论断. 定义

$$q_i(t)=\prod_{\substack{j=1\\ j\neq i}}^{k}(t-\lambda_j)^{n_j},\quad q_i(t)\text{ 的次数 }=n-n_i,$$

由此可见，对所有 $i\neq j$ 有 $q_i(J_{n_j}(\lambda_j))=0$，这是因为 $(J_{n_j}(\lambda_j)-\lambda_jI)^{n_j}=0$. 显然 $q_i(J_{n_i}(\lambda_i))$ 不一定等于 B_i，可是它是非奇异矩阵(因为诸 λ_i 是各不相同的)，且像关于 $J_{n_i}(\lambda_i)$ 的任一个多项式那样，具有形式(3.2.4.3).

因为形如(3.2.4.3)的任一非奇异矩阵的逆具有同一形式，又因为这种形式的任意两个矩阵的乘积还是这同一形式，所以矩阵

$$[q_i(J_{n_i}(\lambda_i))]^{-1}B_i$$

是具有 Toeplitz 形式(3.2.4.3)的上三角矩阵，每个这样的矩阵可以写成关于 $J_{n_i}(\lambda_i)$ 的一个多项式. 例如

$$B_i=b_1^{(i)}(J_{n_i}(\lambda_i)-\lambda_iI)^0+b_2^{(i)}(J_{n_i}(\lambda_i)-\lambda_iI)^1+\cdots+b_{n_i}^{(i)}(J_{n_i}(\lambda_i)-\lambda_iI)^{n_i-1}.$$

因此，存在一个次数至多是 n_i-1 的多项式 $r_i(t)$，使得

$$[q_i(J_{n_i}(\lambda_i))]^{-1}B_i=r_i(J_{n_i}(\lambda_i)).$$

如果现在令 $p_i(t)=q_i(t)r_i(t)$，$p_i(t)$ 的次数至多是 $n-1$.

$$p_i(J_{n_j}(\lambda_j))=q_i(J_{n_j}(\lambda_j))r_i(J_{n_j}(\lambda_j))=0r_i(J_{n_j}(\lambda_j))=0,$$

如果 $i\neq j$，则 $p_i(J_{n_i}(\lambda_i))=q_i(J_{n_i}(\lambda_i))r_i(J_{n_i}(\lambda_i))=B_i$. □

定理的逆命题也同样成立，因而导出非减次矩阵的一个特征：矩阵 $A\in M_n$ 是非减次的，当且仅当每个与 A 可交换的矩阵是 A 的一个多项式.

3.2.5 收敛矩阵 矩阵 $A\in M_n$ 如果有具性质：当 $m\to\infty$ 时，A^m 的所有元素都趋于零，就称 A 为收敛矩阵. 这样的矩阵在数值线性代数的算法分析中起着重要的作用. 如果 A 是对角矩阵，那么显然有：A 收敛，当且仅当 A 的所有特征值的模都严格小于 1，并且同样的结果可以推广到可对角化矩阵.

由于涉及极限运算，如何运用扰动理论把这一结果推广到未必是对角化矩阵的一般情形是不明显的. 但是，可以利用 Jordan 标准形，如果 $A=SJS^{-1}$ 是 A 的 Jordan 标准形，那么 $A^m=SJ^mS^{-1}$，因而，当 $m\to\infty$ 时 $A^m\to 0$ 当且仅当 $m\to\infty$ 时 $J^m\to 0$. 因为 J 是诸 Jordan 块的直和，

137

所以只要考虑 Jordan 块

$$J_k(\lambda)=\begin{bmatrix}\lambda & 1 & & 0\\ & \ddots & \ddots & \\ & & \ddots & 1\\ 0 & & & \lambda\end{bmatrix}=\begin{bmatrix}\lambda & & 0\\ & \ddots & \\ 0 & & \lambda\end{bmatrix}+\begin{bmatrix}0 & 1 & & 0\\ & \ddots & \ddots & \\ & & \ddots & 1\\ 0 & & & 0\end{bmatrix}$$

$$=\lambda I+N_k\in M_k,\quad 其中\ N_k=J_k(0)$$

的各次幂的性质就可以了. 因为，对所有 $m\geqslant k$，$N_k^m=0$，有

$$[J_k(\lambda)]^m=(\lambda I+N_k)^m=\sum_{i=0}^{m}\binom{m}{i}\lambda^i N_k^{m-i}=\sum_{i=m-k+1}^{m}\binom{m}{i}\lambda^i N_k^{m-i}$$

对所有的 $m\geqslant k$ 成立. 因为各对角元都为 λ^m，如果 $J^m\to 0$，则必定有 $\lambda^m\to 0$，这说明 $|\lambda|<1$. 反过来，如果 $|\lambda|<1$. 我们想证明，对每个 $j=0, 1, 2, \cdots, k-1$，

$$当\ m\to\infty\ 时，\ \binom{m}{m-j}\lambda^{m-j}\to 0.$$

然而，

$$\left|\binom{m}{m-j}\lambda^{m-j}\right|=\left|\frac{m(m-1)(m-2)\cdots(m-j+1)\lambda^m}{j!\lambda^j}\right|\leqslant\left|\frac{m^j\lambda^m}{j!\lambda^j}\right|,$$

因此，只要证明当 $m\to\infty$ 时有 $m^j|\lambda|^m\to 0$ 就够了. 看出这个关系的一个简单办法是取对数，并且注意到，因为 $\log|\lambda|<0$，又根据 L'Hopital 法则，当 $x\to\infty$ 时 $(\log x)/x\to 0$，所以当 $m\to\infty$ 时，

$$j\log m+m\log|\lambda|\to-\infty.$$

上述证法主要是利用 A 的 Jordan 标准形来证明，当 $m\to\infty$ 时有 $A^m\to 0$，当且仅当 A 的所有特征值有小于 1 的模. 另一个证明将在(5.6.12)中给出，它根本不依赖 Jordan 标准形.

3.2.6 几何重数-代数重数不等式 矩阵 $A\in M_n$ 的一个特征值的几何重数是 A 的相应于该特征值的 Jordan 块的个数. 这个数小于或等于相应于该特征值的所有 Jordan 块的阶数的和. 这个和是代数重数. 因此一个特征值的几何重数不大于它的代数重数；试与(1.4.9)比较.

138

3.2.7 可对角化矩阵和幂零矩阵 一个矩阵 $A\in M_n$ 是幂零的，是指对某个正整数 k，有 $A^k=0$. 任一 Jordan 块 $J_k(\lambda)$ 可以写成 $J_k(\lambda)=\lambda I+N_k$，其中，$(N_k)^k=0$. 因此，任一 Jordan 块是一个对角矩阵矩与一个幂零矩阵的和.

更一般地，一个 Jordan 矩阵(3.2.1.1)可以写成 $J=D+N$，其中 D 是对角矩阵，它的主对角线与 J 的主对角线相同，且 $N=J-D$. 矩阵 N 是幂零的，且只要 k 是 J 的最大 Jordan 块的阶，就有 $N^k=0$.

最后，如果 $A\in M_n$ 是任一给定的矩阵，且 $A=SJS^{-1}$ 是 Jordan 标准形，那么 $A=SDS^{-1}+SNS^{-1}=A_D+A_N$，其中 A_D 可对角化，A_N 是幂零的，且 $A_DA_N=A_NA_D$，这是因为 D 和 N 都是具有相应的同阶子块的分块对角矩阵，且 D 的诸子块是纯量矩阵.

我们得出，任一 $A\in M_n$ 可以表示成一个可对角化矩阵与一个幂零矩阵的和，且这两个被

加矩阵可交换.

习题

1. 设 $\mathcal{F}=\{A_\alpha:\alpha\in\mathcal{I}\}\subset M_n$ 是给定的矩阵族，且附以指标集 $\mathcal{I}$，又假定存在非减次矩阵 $A_0\in\mathcal{F}$，使得对所有 $\alpha\in\mathcal{I}$，有 $A_\alpha A_0=A_0A_\alpha$. 证明，对每个 $\alpha\in\mathcal{I}$，都存在一个次数至多为 $n-1$ 的多项式 $p_\alpha(t)$ 使得 $A_\alpha=p_\alpha(A_0)$，因而 $\mathcal{F}$ 是一个交换族.

2. 设 $A\in M_n$ 是已知的，λ_i 是 A 的一个特征值. 证明 A 的相应于特征值 λ_i 的最大 Jordan 块的阶数(λ_i 的指标)是使 $\operatorname{rank}(A-\lambda_iI)^k=\operatorname{rank}(A-\lambda_iI)^{k+1}$ 的 $k=1,2,\cdots,n-1$ 的最小值.

3. 如果对某个 $k>n$，$A\in M_n$ 有 $A^k=0$，证明，对某个 $r\leqslant n$，有 $A^r=0$，因此每个幂零矩阵有一个次数不大于其阶数的等于零的幂. **提示**：证明 0 是 A 仅有的特征值. 想一想 A 的 Jordan 标准形是什么样子？再取幂.

4. 设 $J_k(0)$ 是一个已知的 Jordan 块. 试利用(3.2.1)末尾的证法确定矩阵 $J_k^2(0)$ 的二种可能的 Jordan 标准形.

139

5. 设 $A\in M_n$ 是幂零矩阵，因而对某个 k 有 $A^k=0$. 证明，A 的特征多项式是 $p_A(t)=t^n$.

6. 作用在次数至多为 3 的全体多项式所组成的向量空间上的线性变换 $\mathrm{d}/\mathrm{d}t$：$p(t)\to p'(t)$，在基 $\mathcal{B}=\{1,t,t^2,t^3\}$下有基表示

$$\begin{bmatrix}0&1&0&0\\0&0&2&0\\0&0&0&3\\0&0&0&0\end{bmatrix}.$$

这个矩阵的 Jordan 标准形是什么？

7. 使 $A^3=I$ 的矩阵 $A\in M_n$ 的各种可能的 Jordan 形是什么？

8. 具有特征多项式 $p_A(t)=(t+3)^4(t-4)^2$ 的矩阵 $A\in M_6$ 的各种可能的 Jordan 标准形是什么？

9. 试用(3.2.1)中所叙述的方法确定

$$A=\begin{bmatrix}i&1\\1&-i\end{bmatrix}$$

的 Jordan 标准形.

10. 试证定理(3.2.4.2)的证明中的论断：形如(3.2.4.3)的两个矩阵的乘积有同一形式. 再证明一个形如(3.2.4.3)的非奇异矩阵的逆有同一形式. **提示**：A 的逆是一个关于 A 的多项式.

11. 设 $A,B\in M_n$，试用定理(1.3.20)的证明中的恒等式证明，AB 和 BA 的 Jordan 形中的各非奇异子块是恒等的. 这能说明 AB 与 BA 相似吗？如果 AB 与 BA 不相似，讨论这与它们能够相似有多大差别.

12. 假定 $A_1,\cdots,A_k$ 是已知矩阵，其中，$A_i\in M_{n_i}$，$i=1,2,\cdots,k$，又假定 $J_1,\cdots,J_k$ 是它们的相应 Jordan 标准形. 证明直和

$$\begin{bmatrix}A_1&&&0\\&A_2&&\\&&\ddots&\\0&&&A_k\end{bmatrix}\in M_n,\quad n_1+n_2+\cdots+n_k=n$$

（除了差诸对角子块的一个排列以外）有 Jordan 标准形 [140]

$$\begin{bmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_k \end{bmatrix}.$$

13. 设 $A\in M_n$ 和 B，$C\in M_m$ 是已知的. 证明，直和 $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\in M_{n+m}$ 相似于直和 $\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$，当且仅当 B 相似于 C.

14. 设 B，$C\in M_m$ 是已知的，证明，k 次直和

$$\begin{bmatrix} B & & & 0 \\ & B & & \\ & & \ddots & \\ 0 & & & B \end{bmatrix}\in M_{km} \text{ 与 } \begin{bmatrix} C & & & 0 \\ & C & & \\ & & \ddots & \\ 0 & & & C \end{bmatrix}\in M_{km},\quad k\geqslant 1$$

相似，当且仅当 B 与 C 相似.

15. 设 $A\in M_n$ 和 B，$C\in M_m$ 是已知的，证明，直和

$$\begin{bmatrix} A & & & 0 \\ & B & & \\ & & \ddots & \\ 0 & & & B \end{bmatrix}\in M_{n+km} \text{ 与 } \begin{bmatrix} A & & & 0 \\ & C & & \\ & & \ddots & \\ 0 & & & C \end{bmatrix}\in M_{n+km},\quad k\geqslant 1$$

相似，当且仅当 B 与 C 相似.

16. 设 $A\in M_n$ 有 Jordan 标准形 $J_{n_1}(\lambda_1)\oplus\cdots\oplus J_{n_k}(\lambda_k)$. 如果 A 是非奇异形矩阵，证明 A^2 的 Jordan 标准形是 $J_{n_1}(\lambda_1^2)\oplus\cdots\oplus J_{n_k}(\lambda_k^2)$；即 A^2 的 Jordan 标准形恰好由与 A 相同的一组 Jordan 块组成，只是相应的特征值要取平方. 与此类似的结论对所有幂 $A^k(k\geqslant 2)$ 成立吗？举一个 2×2 的例子说明，如果 A 是奇异矩阵，上述结论不成立. **提示**：如果 $A\neq 0$，试证 $J_k^2(\lambda)$ 的 Jordan 标准形（一个 Jordan 块）是 $J_k(\lambda^2)$. 再证明，如果 $\lambda\neq 0$，

$$\operatorname{rank}[J_k(\lambda)-\lambda I]^m = \operatorname{rank}[J_k^2(\lambda)-\lambda^2 I]^m,\quad m=1,2,\cdots,k$$

17. 如果 $A\in M_n$，证明，$\operatorname{rank} A=\operatorname{rank} A^2$，当且仅当特征值 $\lambda=0$ 的几何重数与代数重数相等，即 A 的 Jordan 标准形中相应于 $\lambda=0$ 的所有 Jordan 块（如果有的话）是 1×1 的. [141]

进一步阅读 一个给定的矩阵与其转置间的相似变换矩阵总可以取为对称矩阵. 该事实的证明可参看 O. Taussky and H. Zassenhaus, "On the Similarity Transformation between a Matrix and Its Transpose," *Pacific J. Math.* 9(1959), 893-896. 在(3.2.4)节末提到了定理(3.2.4.2)的逆命题也成立，其证明可见[HJ].

3.3 多项式和矩阵：极小多项式

如果 $p(t)=t^k+a_{k-1}t^{k-1}+a_{k-2}t^{k-2}+\cdots+a_1t+a_0$ 是给定的多项式，那么，对于任一 $A\in M_n$，可以定义

$$p(A) \equiv A^k + a_{k-1}A^{k-1} + a_{k-2}A^{k-2} + \cdots + a_1 A + a_0 I.$$

在多项式与矩阵之间存在一个重要的相互关系. 我们已经论述了特征多项式的重要作用，但是还有其他一些与一个方阵相关联的多项式，其中之一是极小多项式.

Cayley-Hamilton 定理(2.4.2)确认，对每一个 $A\in M_n$，存在一个 n 次多项式(特征多项式)$p_A(t)$，使得 $p_A(t)=0$. 说一个多项式零化 A，是指它在 A 的值是 0 矩阵，也可能有零化 A 的 $n-1$ 次多项式，或 $n-2$ 次多项式，但是很明显，因为只存在有限多种可能性，所以，对每个 $A\in M_n$，有一个零化 A 的次数最小的多项式，且它的最小次数至多是 n. 如果 $p(A)=0$，那么对任一 $c\in\mathbf{C}$，$cp(A)=0$，显然，总可以规范化一个非零的零化多项式，使其最高次项的系数是+1. 如果一个多项式的最高次数有系数 1，就说它是首一多项式，应指出的是，一个首一多项式不可能恒等于零.

3.3.1 定理 $A\in M_n$ 是已知矩阵. 则仅有一个零化 A 的次数最低的首一多项式 $q_A(t)$. 这个多项式的次数至多是 n. 如果 $p(t)$ 是使 $p(A)=0$ 的任一多项式，那么，$q_A(t)$ 必除尽 $p(t)$.

证明：特征多项式是零化 A 的一个例子. 它是 n 次多项式，于是存在次数为 $m\leqslant n$ 的首一

[142] 多项式 $q(t)$，使得 $q(A)=0$. 如果 $p(t)$ 零化 A，且 $q(t)$ 是零化 A 的次数最低的首一多项式，那么 $q(t)$ 的次数必须小于或等于 $p(t)$ 的次数. 因此，根据 Euclid 算法，存在多项式 $h(t)$ 以及次数小于 $q(t)$ 的次数的多项式 $r(t)$，使得 $p(t)=q(t)h(t)+r(t)$. 但是 $0=p(A)=q(A)h(A)+r(A)=0h(A)+r(A)$，因而 $r(A)=0$. 如果 $r(t)\neq 0$，可以把它规范化，从而得到一个零化 A 的而次数小于 $p(t)$ 的次数的首一多项式. 因为这与 $p(t)$ 的极小性相矛盾，所以得出 $r(t)=0$，因而 $q(t)$ 除尽 $p(t)$，且有商 $h(t)$. 如果有两个零化 A 的次数最低的首一多项式，上述论证说明，每一个都除尽另一个；因为它们的次数相同，所以其中一个必须是另一个的一个纯量倍数. 然而，它们都是首一的，纯量因子必须是+1，因而它们是恒等的. □

3.3.2 定义 设 $A\in M_n$ 是已知矩阵. 零化 A 的，次数最小的唯一首一多项式 $q_A(t)$ 称为 A 的极小多项式.

3.3.3 推论 相似的矩阵有相同的极小多项式.

证明：如果 A，B，$S\in M_n$，且 $A=SBS^{-1}$，那么，$q_B(A)=q_B(SBS^{-1})=Sq_B(B)S^{-1}=0$ 于是 $q_B(t)$ 的次数不小于 $q_A(t)$ 的次数. 可是 $B=S^{-1}AS$，所以同样的论证说明 $q_A(t)$ 的次数不小于 $q_B(t)$ 的次数. 因此，这两个首一多项式有相同的最低次数且都零化 A，于是根据定理(3.3.1)，它们必须恒等. □

3.3.4 推论 对于每个 $A\in M_n$，极小多项式 $q_A(t)$ 除尽特征多项式 $p_A(t)$ 此外，$q_A(\lambda)=0$，当且仅当 λ 是 A 的特征值，因而 $p_A(t)=0$ 的每个根是 $q_A(t)=0$ 的根.

证明：因为 $p_A(t)=0$，由定理可知，存在多项式 $h(t)$，使得 $p_A(t)=h(t)q_A(t)$. 这个分解使我们看到，$q_A(t)=0$ 的每个根是 $p_A(t)=0$ 的根，因而 $q_A(t)=0$ 的每个根是 A 的特征值. 如果 λ 是 A 的特征值，且 $x\neq 0$ 是相应的特征向量，那么 $Ax=\lambda x$ 且 $0=q_A(A)x=q_A(\lambda)x$，因此 $q_A(\lambda)=0$. □

[143] 这后一个推论说明，如果把特征多项式 $p_A(t)$ 完全分解成

$$p_A(t) = \prod_{i=1}^{m}(t-\lambda_i)^{s_i}, \quad 1 \leqslant s_i \leqslant n, \quad s_1+s_2+\cdots+s_m = n, \tag{3.3.5a}$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 各不相同，那么极小多项式 $q_A(t)$ 必须有形式

$$q_A(t) = \prod_{i=1}^{m}(t-\lambda_i)^{r_i}, \quad 1 \leqslant r_i \leqslant s_i. \tag{3.3.5b}$$

原则上，这给出了一个求已知矩阵 A 的极小多项式的算法：

1. 首先，假定通过求特征多项式以及对它做完全分解，计算出 A 的各个特征值和它们的代数重数．设法确定分解(3.3.5a)．

2. 在(3.3.5b)中，有有限多个乘积形式的多项式，从所有 $r_i=1$ 的乘积开始，通过直接验算，确定零化 A 的，次数最低的多项式．这就是所求的极小多项式．

从数值计算考虑，这不是一个好的算法．因为对于一个大矩阵，要牵涉到分解其特征多项式的问题．但是，对于手算一些形式简单的小矩阵的极小多项式，它可能是很有效的．

练习 设 $A=\begin{bmatrix}\lambda & a\\ 0 & \lambda\end{bmatrix}$，求 A 的极小多项式 $q_A(t)$．

A 的 Jordan 标准形和 A 的极小多项式之间有着密切的联系．假定 $A=SJS^{-1}$ 是 A 的 Jordan 标准形，且先假定

$$J = \begin{bmatrix}\lambda & 1 & & 0\\ & \ddots & \ddots & \\ & & \ddots & 1\\ 0 & & & \lambda\end{bmatrix} \in M_n$$

只是一个 Jordan 块．A 的特征多项式是 $(t-\lambda)^n$，又因为如果 $k<n$，则 $(J-\lambda I)^k \neq 0$，所以极小多项式也是 $(t-\lambda)^n$．如果

$$J = \begin{bmatrix}J_{n_1}(\lambda) & 0\\ 0 & J_{n_2}(\lambda)\end{bmatrix} \in M_n,$$

其中 $n_1 \geqslant n_2$，那么 J 的特征多项式仍为 $(t-\lambda)^n$，可是现在 $(J-\lambda I)^{n_1}=0$，并且没有较低的幂等于零．因此，极小多项式是 $(t-\lambda)^{n_1}$．如果有更多的子块，结果是相同的：J 的极小多项式是 $(t-\lambda)^r$，其中 r 是相应于 λ 的最大 Jordan 块的阶数．如果 J 是一般的 Jordan 矩阵，那么极小多项式必定包含每个不同特征值 λ_i 的一个因式 $(t-\lambda_i)^{r_i}$，且 r_i 一定是相应于 λ_i 的最大 Jordan 块的阶数；没有更低的幂可以零化相应于 λ_i 的所有 Jordan 块，并且也不需要更高的幂．因为相似的矩阵有相同的极小多项式，我们已经证明了下述定理：

3.3.6 定理 设 $A \in M_n$ 是已知矩阵，其不同的特征值是 $\lambda_1, \cdots, \lambda_m$．$A$ 的极小多项式是

$$q_A(t) = \prod_{i=1}^{m}(t-\lambda_i)^{r_i}, \tag{3.3.7}$$

其中 r_i 是 A 的相应于特征值 λ_i 的最大 Jordan 块的阶数．

实际上，这个结果在计算极小多项式时不是很有用的，因为确定一个矩阵的 Jordan 标准形比确定它的极小多项式更难一些；况且，只要知道了一个矩阵的诸特征值，它的极小多项式

可以通过简单的试凑法来确定. 不过，有一个重要的理论性推论. 因为一个矩阵可对角化，当且仅当它的所有 Jordan 块有阶数 1，所以，在(3.3.7)中，可对角化的一个必要充分条件是所有 $r_i=1$.

3.3.8 推论 设 A 是已知矩阵，它的不同的特征值是 $\lambda_1, \lambda_2, \cdots, \lambda_m$. 那么，$A$ 可对角化，当且仅当 $q(A)=0$，其中

$$q(t)=(t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_m). \tag{3.3.9}$$

在确定一个已知矩阵是否可对角化时，这个准则是很实用的，因为如果知道一个已知矩阵的特征值，就容易写出多项式(3.3.9). 然后看它是否零化 A，如果它零化 A，它一定是 A 的极小多项式. 因为没有更低次的多项式能以 A 的 m 个不同的特征值为根. 把上述结果用下述几种等价的方式来表述，有时很有用.

3.3.10 推论 设 $A\in M_n$ 是已知矩阵. 下列条件中的每一个都是 A 可对角化的必要充分条件：

(a) 极小多项式 $q_A(t)$ 有不同的线性因子.

(b) $q_A(t)$ 的每个根有重数 1.

145 (c) 对于使 $q_A(t)=0$ 的所有根，导数 $q'_A(t)\neq 0$.

对于已知矩阵 $A\in M_n$，已经讨论了求一个零化 A 的，次数最低的首一多项式的问题. 然而，关于它的逆问题又如何呢? 给定一个首一多项式

$$p(t)=t^n+a_{n-1}t^{n-1}+a_{n-2}t^{n-2}+\cdots+a_1t+a_0, \tag{3.3.11}$$

存在一个矩阵 A 使得 $p(t)$ 是其极小多项式吗? 如果有，A 的阶至少必须为 $n\times n$；要求这样一个矩阵是不困难的. 考察矩阵

$$A=\begin{bmatrix} 0 & & & & -a_0 \\ 1 & 0 & & 0 & \vdots \\ & 1 & \ddots & & \vdots \\ 0 & & \ddots & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{bmatrix}\in M_n, \tag{3.3.12}$$

并且看到

$$\begin{aligned}
Ie_1 &= e_1 = A^0e_1,\\
Ae_1 &= e_2 = Ae_1,\\
Ae_2 &= e_3 = A^2e_1,\\
Ae_3 &= e_4 = A^3e_1,\\
&\ \ \vdots \qquad \vdots \\
Ae_{n-1} &= e_n = A^{n-1}e_1,\\
Ae_n &= -a_{n-1}e_n-a_{n-2}e_{n-1}-\cdots-a_1e_2-a_0e_1\\
&= -a_{n-1}A^{n-1}e_1-a_{n-2}A^{n-2}e_1-\cdots-a_1Ae_1-a_0e_1=A^ne_1\\
&= [A^n-p(A)]e_1.
\end{aligned}$$

因而

$$p(A)e_1=(a_0e_1+a_1Ae_1+a_2A^2e_1+\cdots+a_{n-1}A^{n-1}e_1)+A^ne_1$$
$$=[p(A)-A^n]e_1+[A^n-p(A)]e_1=0.$$

另外，对每个 $k=1, 2, \cdots, n$，$p(A)e_k=p(A)A^{k-1}e_1=A^{k-1}p(A)e_1=A^{k-1}0=0$. 因为对每个基向量 e_k 有 $p(A)e_k=0$，我们得知 $p(A)=0$. 因此，$p(t)$是零化 A 的 n 次首一多项式. 如果存在零化 A 的，且有较低次数 $m>n$ 的多项式 $q(t)=t^m+b_{m-1}t^{m-1}+\cdots+b_1t+b_0$，那么

$$0=q(A)e_1=A^me_1+b_{m-1}A^{m-1}e_1+\cdots+b_1Ae_1+b_0e_1$$
$$=e_{m+1}+b_{m-1}e_m+\cdots+b_1e_2+b_0e_1=0,$$

[116]

它推出基向量 e_{m+1} 与基向量 $e_1, e_2, \cdots, e_m$ 线性相关. 因为这是不可能的，由此得出 $p(t)$是零化 A 的，次数最低的唯一首一多项式. 另外，因为 $p(t)$有次数 n，$A\in M_n$，且特征多项式 $p_A(t)$也是零化 A 的 n 次首一多项式，所以(3.3.11)必须是(3.3.12)的特征多项式.

3.3.13 定义　矩阵(3.3.12)称为多项式(3.3.11)的友矩阵.

已经证明了下述定理：

3.3.14 定理　每个首一多项式既是它的友矩阵的极小多项式，又是其特征多项式.

稍后，我们将提出几种不同的方法来确定包含一个矩阵的诸特征值的区域. 因为一个多项式的零点是其友矩阵的特征值，所以这些方法可以用来估计一个多项式的各零点. 见(5.6)节.

如果 $A\in M_n$ 是一个已知矩阵，我们可以计算其特征多项式 $p_A(t)$以及多项式 $p_A(t)$的友矩阵(3.3.12). 如果 A 相似于这个友矩阵，那么(因为相似的矩阵有相同的极小多项式)，由(3.3.14)可知，A 的极小多项式 $q_A(t)$一定恒等于 A 的特征多项式 $p_A(t)$. 一般说来不会有这种情况，不过，如果 $A\in M_n$ 是一个其极小多项式 $q_A(t)$与其特征多项式 $p_A(t)$恒等的矩阵，那么 A 的 Jordan 标准形一定恰好含有每个不同的特征值的一个 Jordan 块. 每个 Jordan 块的阶数等于作为 A 的特征(极小)多项式的一个零点的相应特征值的重数. 但是，多项式 $p_A(t)$的友矩阵的 Jordan 标准形与 A 恰好有相同的 Jordan 块结构. 因而它必须与 A 相似，这番论证就是下述定理的一个证明.

3.3.15 定理　矩阵 $A\in M_n$ 相似于其特征多项式的友矩阵，当且仅当 A 的极小多项式与特征多项式恒等.

练习　证明，$A\in M_n$ 相似于其特征多项式的友矩阵，当且仅当 A 是非减次矩阵.

[147]

习题

1. 设 $A, B\in M_3$ 是幂零矩阵，证明，A 与 B 相似，当且仅当 A 和 B 有相同的极小多项式，这在 M_4 中成立吗?

2. 假定 $A\in M_n$ 是已知的，且 A 的不同的特征值 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 也是已知的. 试用(3.3.6)证明，极小多项式(3.3.7)可以用下述算法来确定：对每个 $i=1, 2, \cdots, m$，计算 $(A-\lambda_iI)^k, k=1, 2, \cdots, n$. 设 r_i 是使 $\operatorname{rank}(A-\lambda_iI)^k=\operatorname{rank}(A-\lambda_iI)^{k+1}$ 的最小 k 值，这个数 r_i 称为特征值 λ_i 的指标.

3. 矩阵 $A\in M_n$ 是幂等的，如果 $A^2=A$. 利用(3.3.10)证明，每个幂等矩阵是可对角化的. **提示**：试证 $t^2-t=t(t-1)$零化 A. A 的极小多项式是什么? 如果 A 是三次幂等的($A^3=$

A)，你能说些什么？如果 $A^k=A$ 呢？

4. 如果 $A\in M_n$，且对某个 $k<n$，有 $A^k=0$，证明对某个 $r\leqslant n$ 有 $A^r=0$. 因此，每个幂零矩阵有一个其指数不大于该矩阵的阶数的等于零的幂. **提示**：如果 $p(t)=t^k$ 零化 A，想一想(3.3.1)关于极小多项式是怎么说的？

5. 证明 Gram-Schmidt 过程有下述应用，那就是，对于一个给定的矩阵 $A\in M_n$，既不需要知道它的特征多项式又不需要知道它的任何特征值，也可以直接计算 A 的极小多项式.

(a)设映射 T：$M_n\to \mathbf{C}^{n^2}$ 定义如下：对于任一 $A\in M_n$，把它按列块分成 $A=[a_1, a_2, \cdots, a_n]$；设 $T(A)$表示 $\mathbf{C}^{n^2}$ 中唯一确定的向量，它的前 n 个分量是第 1 列 a_1 的各个分量，它的 $n+1$ 到 $2n$ 个分量是第 2 列的各个分量，如此等等. 证明，这个映射 T 是向量空间 M_n 与 $\mathbf{C}^{n^2}$ 间的同构(线性的、一对一的和到上的).

(b)考察 $\mathbf{C}^{n^2}$ 中的诸向量

$$v_0=T(I), v_1=T(A), v_2=T(A^2), \cdots, v_k=T(A^k), \cdots,$$

其中 $k=0, 1, 2, \cdots, n$. 试用 Cayley-Hamilton 定理证明，$\{v_0, v_1, \cdots, v_n\}$是一个相关组.

(c)把 Gram-Schmidt 过程按给定的顺序应用于向量组$\{v_0, v_1, \cdots, v_n\}$直到产生第一个零向量为止. 为什么一定会得到一个零向量？

(d)如果 Gram-Schmidt 过程在第 k 步产生第一个零向量，证明 $k-1$ 是 A 的极小多项式的次数.

148

(e)如果 Gram-Schmidt 过程产生向量 $\alpha_0 v_0+\alpha_1 v_1+\cdots+\alpha_{k-1}v_{k-1}=0$，证明

$$T^{-1}(\alpha_0 v_0+\alpha_1 v_1+\cdots+\alpha_{k-1}v_{k-1})=\alpha_0 I+\alpha_1 A+\alpha_2 A^2+\cdots+\alpha_{k-1}A^{k-1}=0,$$

由此得出，$q_A(t)=(\alpha_{k-1}t^{k-1}+\cdots+\alpha_2 t^2+\alpha_1 t+\alpha_0)/\alpha_{k-1}$是 A 的极小多项式. 为什么 $\alpha_{k-1}\neq 0$？

6. 按习题 5 中算法的要求进行计算，分别求$\begin{bmatrix}1&1\\0&2\end{bmatrix}$，$\begin{bmatrix}1&1\\0&1\end{bmatrix}$和$\begin{bmatrix}1&0\\0&1\end{bmatrix}$的极小多项式.

7. 考察 $A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$和 $B=\begin{bmatrix}0&0\\0&1\end{bmatrix}$，说明 AB 和 BA 的极小多项式不一定相同. 但 AB 和 BA 的特征多项式是相同的. 说明特征多项式和极小多项式之间为什么存在这一差别.

8. 设 $A_i\in M_{n_i}$，$i=1, 2, \cdots, k$，且设 $q_i(t)$表示每个 A_i 的极小多项式. 证明直和

$$A=\begin{bmatrix}A_1 & & & 0\\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_k\end{bmatrix}$$

的极小多项式是 $q_1(t)$，$q_2(t)$，$\cdots$，$q_k(t)$的最小公倍式. 这个唯一的，次数最低的首一多项式可被每个 $q_i(t)$除尽. 注意，上述论证给出了引理(1.3.10)的一个不同的证明.

9. 如果 $A\in M_5$ 的特征多项式 $p_A(t)=(t-4)^3(t+6)^2$ 和极小多项式 $q_A(t)=(t-4)^2(t+6)$，那么 A 的 Jordan 标准形是什么？

10. 用直接计算证明，多项式(3.3.11)是友矩阵(3.3.12)的特征多项式. **提示**：利用余子式计算行列式.

11. 有时，多项式(3.3.11)的友矩阵定义为

$$\begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_0 \\ 1 & 0 & \cdots & 0 \\ & 1 & \ddots & \\ 0 & & \ddots & \\ & & 1 & 0 \end{bmatrix} \quad 或 \quad \begin{bmatrix} 0 & 1 & & 0 \\ \vdots & \ddots & \ddots & \\ 0 & & 0 & 1 \\ -a_n & -a_1 & \cdots & -a_{n-1} \end{bmatrix}$$

[149]

证明，这两个矩阵与(3.3.12)有相同的性质：(3.3.11)既是该矩阵的极小多项式，又是它的特征多项式.

12. 说明，不存在其极小多项式为 x^2+1 的 3×3 实矩阵，但是存在具有这一性质的一个 2×2 实矩阵及一个 3×3 复矩阵. **提示**：利用(3.3.4).

13. 虽然相似的矩阵有相同的特征多项式和极小多项式，说明 4 阶或更高阶的两个矩阵可能有相同的极小多项式和特征多项式而它们不相似. **提示**：考虑

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad 和 \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

说明 4 是这种情况可能出现的最低阶数.

14. 如果 A，$B \in M_n$ 是相似的，而 $p(t)$ 是一个多项式，那么 $p(A)=0$ 当且仅当 $p(B)=0$. 试用前一个习题的例子说明，即使 A 与 B 不相似，$p(A)=0$ 当且仅当 $p(B)=0$ 也可能对每个多项式 $p(t)$ 成立. 为什么会出现这种情形？

15. 设 $A \in M_n$ 是已知矩阵，且设 $P(A)=\{p(A):\ p(t)是一个多项式\}$. 说明 $P(A)$ 是 M_n 的子空间，且它还是 M_n 的子代数[$P(A)$在乘法下封闭]. 证明 $P(A)$ 的维数是 A 的极小多项式的次数.

16. 如果 A，$B \in M_n$ 有相同的特征多项式和相同的极小多项式，且它们的极小多项式与它们的特征多项式相同，证明 A 与 B 相似. 再利用这个事实证明，关于习题 9 中所提到的友矩阵的其他各种形式都相似于(3.3.12).

3.4 其他标准形和分解

除了 Jordan 标准形以外，还有几种其他的矩阵分解，它们可以用于各种不同的情形.

当矩阵 A 只有实元素时，首先考虑 Jordan 标准形(3.1.12)的一个变形. 这时，所有的非实特征值必须成共轭对出现. 此外，如果 A 是实的，那么 $\operatorname{rank}(A-\lambda I)^k=\operatorname{rank}(\overline{A}-\lambda I)^k=\operatorname{rank}(A-\bar{\lambda} I)^k$ 对所有 $\lambda \in \mathbf{C}$ 和所有 $k=1,2,\cdots$ 成立，因而，相应于任一特征值 λ 的诸 Jordan 块的结构与相应共轭特征值 $\bar{\lambda}$ 的诸 Jordan 块的结构相同. 因此，相应于各非实特征值的所有阶数的全体 Jordan 块(不仅仅是 1×1 子块)成相同阶数的共轭对出现.

[150]

例如，如果 λ 是实矩阵 A 的一个非实特征值，且 $J_2(\lambda)$ 以某个确定的重数出现在 A 的 Jordan 标准形中，那么 $J_2(\bar{\lambda})$ 也必须以相同的重数出现. 分块矩阵

$$\begin{bmatrix} J_2(\lambda) & 0 \\ 0 & J_2(\bar{\lambda}) \end{bmatrix} = \left[\begin{array}{cc:cc} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ \hdashline 0 & 0 & \bar{\lambda} & 1 \\ 0 & 0 & 0 & \bar{\lambda} \end{array}\right] \tag{3.4.1}$$

(交换其第 2，第 3 行和列)置换相似于

$$\left[\begin{array}{cc:cc} \lambda & 0 & 1 & 0 \\ 0 & \bar{\lambda} & 0 & 1 \\ \hdashline 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \bar{\lambda} \end{array}\right] = \begin{bmatrix} D(\lambda) & I \\ 0 & D(\lambda) \end{bmatrix},$$

其中

$$D(\lambda) \equiv \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix} \in M_2, \text{且 } I \in M_2.$$

一般地，任一形如

$$\begin{bmatrix} J_k(\lambda) & 0 \\ 0 & J_k(\bar{\lambda}) \end{bmatrix} \in M_{2k} \tag{3.4.2}$$

的矩阵都置换相似于分块矩阵

$$\begin{bmatrix} D(\lambda) & I & & 0 \\ & D(\lambda) & I & \\ & & \ddots & \\ 0 & & \ddots & I \\ & & & D(\lambda) \end{bmatrix} \in M_{2k},$$

其中主对角线上有 k 个子块 $D(\lambda)$，而在上对角线上有 $k-1$ 个 2×2 单位矩阵

151

$$SD(\lambda)S^{-1} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \equiv C(a,b), \tag{3.4.3}$$

其中 $\lambda = a + ib$，a，$b \in \mathbf{R}$，且 $S = \begin{bmatrix} -i & i \\ 1 & -1 \end{bmatrix}$，因此，每个具有非实 λ 的共轭 2×2 Jordan 块的子块(3.4.1)经 $\begin{bmatrix} S & 0 \\ 0 & S \end{bmatrix}$ 相似于一个具有形状

$$\left[\begin{array}{cc:cc} a & b & 1 & 0 \\ -b & a & 0 & 1 \\ \hdashline 0 & 0 & a & b \\ 0 & 0 & -b & a \end{array}\right] = \begin{bmatrix} C(a,b) & I \\ 0 & C(a,b) \end{bmatrix}$$

的 4×4 实分块矩阵．一般地，每个具有非实 λ 的共轭 $k\times k$ Jordan 块子块对(3.4.2)相似于具有形状

$$C_k(a,b) = \begin{bmatrix} C(a,b) & I & & 0 \\ & C(a,b) & \ddots & \\ & & \ddots & I \\ 0 & & & C(a,b) \end{bmatrix} \tag{3.4.4}$$

的 $2k\times 2k$ 实分块矩阵. 这些论述导出了实 Jordan 标准形.

3.4.5 定理 每个实矩阵 $A\in M_n(R)$ 相似于具有形状

$$\begin{bmatrix} C_{n_1}(a_1,b_1) & & & & & & 0 \\ & C_{n_2}(a_2,b_2) & & & & & \\ & & \ddots & & & & \\ & & & C_{n_p}(a_p,b_p) & & & \\ & & & & J_{n_q}(\lambda_q) & & \\ & & & & & \ddots & \\ 0 & & & & & & J_{n_r}(\lambda_r) \end{bmatrix} \quad (3.4.6)$$

的分块对角实矩阵，其中，对于 $k=1, 2, \cdots$，$\lambda_k=a_k+ib_k$ 是非实特征值，a_k 和 b_k 是实数，而 λ_q，…，λ_r 是 A 的实特征值. 每个实分块三角矩阵 $C_{n_k}(a_k, b_k)\in M_{2n_k}$ 具有形状(3.4.4)，且对应于 A 的 Jordan 标准形(3.1.12)中具有非实 λ_k 的一对共轭 Jordan 块 $J_{n_k}(\lambda_k)$，$J_{n_k}(\bar{\lambda}_k)\in M_{n_k}$. (3.4.6)中的实 Jordan 块 $J_{n_k}(\lambda_k)$ 恰好是(3.1.12)中具有实 λ_k 的 Jordan 块.

我们已经从一个实矩阵的一般(复)Jordan 标准形(3.1.12)推导出它的实 Jordan 标准形. 这样处理的优点是，它正好说明实子块 $C_{n_k}(a_k, b_k)$ 的阶数和个数与 A 的复 Jordan 块结构是什么关系. 但是这样处理的缺点是，把 A 变换成(3.4.6)的相似矩阵可以选取实矩阵这个事实是不明显的. [152]

事实上，如果 A 是实矩阵，那么总有一个非奇异的实矩阵 S，使得 $S^{-1}AS$ 成为实 Jordan 形(3.4.6). 可以通过下述三个步骤来证明这一事实，这些步骤正是在(3.1)中的 Jordan 标准形定理的证明中所采用的. 首先，用 Schur 三角化定理的实形式(2.3.4)代替复形式(2.3.1). 在第二步和第三步，可以模仿在复情形下的论述，说明在每个步骤都可以采用实相似，最终将实形式(2.3.4)化简成变了形的诸三角子块或诸 Jordan 对角子块的一个直和，其中，主对角线上可能有形如(3.4.3)的 2×2 实子块 $C(a, b)$.

复 Jordan 标准形(3.1.12)是诸上三角矩阵的一个直和，而实 Jordan 形(3.4.6)是诸 Hessenberg 矩阵或"近乎上三角"矩阵的一个直和，这是因为每个实 2×2 子块 $C(a, b)$ 在主对角线下有一个元素.

我们也能提出其他的标准形，它们是一些友矩阵的直和. 这种形式对复矩阵有一定的要求，但也有以下优点，它们适用于不同于 **C** 的域，而 Jordan 标准形对这样的域却无能为力.

设 $A\in M_n$ 是一个已知矩阵，且它的 Jordan 标准形是(3.1.12). 把相应于每个不同的特征值的所有 Jordan 块组合在一起. 从每一组中选一个最大阶数的 Jordan 块，并把它从该组中取出. 设 B_1 表示所有这些取出的子块的直和. A 有多少不同的特征值，那么 B_1 中将有多少个直加矩阵. 现在从每一组所余下的子块中选一个最大阶数的 Jordan 块，且把它从该组中取出. 设 B_2 表示所有这些子块的直和. B_2 中直加矩阵的个数可能少于 B_1，因为某个子块可能已是空集；即 A 的某个特征值可能只有一个相应于它的 Jordan 块. 继续作直和 B_1，B_2，B_3，…，B_s，直到所有 Jordan 块组都成为空集. B_k 的阶数是单调不增的. 于是 $B_1\oplus B_2\oplus\cdots\oplus B_s$ 置换

相似于 A 的原 Jordan 形(3.1.12).

由于已作出的各个直和 B_k 是按上述方式构造出来的，所以每个 B_k 的极小多项式与特征多项式是相同的. 事实上，B_k 的特征(极小)多项式恰好是 A 的极小多项式. 因此，根据(3.3.15)，每个 B_k 相似于它的特征(极小)多项式的友矩阵.

[153] 矩阵 B_k 的特征(极小)多项式称为 A 的不变因式 $f_k(t)$. 注意，它们的次数是单调不增的，且每个 $f_{k+1}(t)$ 除尽 $f_k(t)$，$k=1, 2, \cdots, s-1$. 第一个不变因式 $f_1(t)=q_{B_1}(t)$ 是 A 的极小多项式，且所有不变因式的乘积是 A 的特征多项式. 诸不变因式按一种确定的方式由 A 的 Jordan 块结构所确定，而 A 的 Jordan 块结构被 A 的各特征值 λ_i 及 $(A-\lambda_i I)$ 的幂的秩序列所确定. 因此，相似的矩阵将有相同的不变因式，又因为不变因式也确定 A 的 Jordan 块结构，所以，具有相同的不变因式的两个矩阵必定相似. 因此一个矩阵的不变因式序列(它包括极小多项式且确定特征多项式)是多项式相似不变量完全集：两个矩阵 $A, B\in M_n$ 相似，当且仅当它们的各个不变因式是恒等的.

另一种描述 A 的不变因式的方式是定义 $f_1(t)=(t-\lambda_1)^{r_1}\cdots(t-\lambda_m)^{r_m}$ 为 A 的极小多项式. 首先 A 的 Jordan 标准形中去掉相应于 $f_1(t)$ 的每个因式 $(t-\lambda_i)^{r_i}$ 的一个 Jordan 块(这些正是构成 B_1 的诸 Jordan 块)，且设 $f_2(t)=(t-\lambda_1)^{s_1}\cdots(t-\lambda_m)^{s_m}$ 是余下的 Jordan 形的极小多项式. 之后又去掉相应于每个因式 $(t-\lambda_i)^{s_i}$ 的一个子块，且设 $f_3(t)$ 是余下的 Jordan 形的极小多项式，等等. 诸不变因式 $f_k(t)$ 正好是一系列被压缩的矩阵的极小多项式，在每一步都要从被压缩的矩阵中去掉某些 Jordan 块.

用不变因式来刻划相似矩阵在理论上是受欢迎的，这是因为它清楚地说明为什么极小多项式和特征多项式一般不足以判别相似性，但是它实际上没有把任何内容加到已经知道的准则里：两个矩阵相似，当且仅当它们的 Jordan 标准形是相同的.

另一方面，这样的描述可导出 A 的一个新的标准形，称之为有理形，这是因为只需对矩阵 A 的诸元素施以有理运算便可算出它的各个不变因式.

3.4.7 定理 每个矩阵 $A\in M_n$ 相似于它的各不变因式的友矩阵的直和.

因为已经有了 Jordan 标准形，复矩阵的有理形(3.4.7)似乎就没有什么用处了. 引进了它的理由是，它的另一种形式在任意域上成立，不只限于复数域. 域 $\mathbf{F}$ 上的任一矩阵在 $\mathbf{F}$ 上相似于它的各不变因式的友矩阵的直和，这些不变因式是系数取自 $\mathbf{F}$ 的唯一确定的多项式. 我们对

[154] 实数域来说明这一事实.

如果 $A\in M_n(\mathbf{R})$ 是已知实矩阵，那么 A(可能经一个复相似变换)相似于直和 $B_1\oplus B_2\oplus\cdots\oplus B_s$，按这个次序，它相似于它的不变因式(项 B_k 的特征多项式，$k=1, \cdots, s$)的友矩阵的一个直和. A 的相应于非实特征值的诸 Jordan 块必须成共轭对出现. 如果相应于一个非实特征值的一个子块出现在任一 B_k 中，那么它的共轭也一个出现在同一个 B_k 中，$k=1, \cdots, s$. 于是，每个 B_k 有一个实特征多项式，且由(3.4.7)所确定的形式是一个实矩阵. 实矩阵有理形的这个形式实际上可经一个实相似得到，我们略去了它的证明.

3.4.8 定理 每个实矩阵 $A\in M_n(\mathbf{R})$ 在 $\mathbf{R}$ 上相似于实首一多项式 $p_1(t), p_2(t)\cdots p_s(t)$ 的友矩阵的一个直和，其中每个 $p_{k+1}(t)$ 除尽 $p_k(t)$，$k=1, 2, \cdots, s-1$. 多项式 $p_1(t)$ 是 A 在 $\mathbf{R}$ 上的极

小多项式，乘积 $p_1(t)\cdots p_r(t)$是 A 的特征多项式，且每个 $p_k(t)$是 A 在 **R** 上的一个不变因式. 多项式 $p_k(t)$是唯一确定的，因而 **R** 上的两个实矩阵相似，当且仅当它们有相同的不变因式.

我们着重指出，相同形式的定理在有理数域 **Q** 或任一其他域上也成立. 之所以得有理形这个名称，其原因不外乎是，将矩阵 $A\in M_n(\mathbf{F})$化成所述形式一般可以对 A 在 **F** 中的元素作有限多次有理计算来完成. 因此，如果 **F** 是有理数域，只需采用具有有理元素的相似矩阵和有理系数多项式.

一个涉及友矩阵的不同标准形也可以由 Jordan 标准形(3.1.12)推导出来. 注意到每个单独的 Jordan 块有性质：它的极小多项式和特征多项式是相同的. 于是每个 Jordan 块 $J_{n_i}(\lambda_i)$相似于它的特征多项式$(t-\lambda_i)^{n_i}$的友矩阵. 因此，整个 Jordan 标准形相似于这些多项式$(t-\lambda_i)^{n_i}$的友矩阵的一个直和；这些多项式称为 A 的*初等因子*. 值得指出的是，用这种方式分解 A 时的直加矩阵的个数比分解 A 为有理形时要多；每个不变因式可能提供多个初等因子. A 的所有初等因子的乘积是 A 的特征多项式.

如果 $A\in M_n(\mathbf{R})$，则在 **C** 上计算它的 Jordan 形和初等因子时要注意到它们必须成共轭对出现. 如果将子块 $J_{n_i}(\lambda)$和 $J_{n_i}(\bar{\lambda})$组合成一个直和，所得到的子块有实多项式$(t-\lambda)^{n_i}(t-\bar{\lambda})^{n_i}$作为它的特征多项式和极小多项式，因而它相似于$[t^2-(2\mathrm{Re}\lambda)t+|\lambda|^2]^{n_i}$的实友矩阵. 后一个多项式是 A 的一个实初等因子. 实线性因子的幂也可以作为 A 的每个实特征值的初等因子而出现.

为了避免混淆，通常将与初等因子相关联的标准形称为*有理标准形*.

3.4.9 定理 每个矩阵 $A\in M_n(\mathbf{R})$在 **R** 上相似于它的各(实)初等因子的友矩阵的直和.

相同形式的结论在任一域 **F** 上也成立：每个 $A\in M_n(\mathbf{F})$在 **F** 上也相似于它的各初等因子的友矩阵的直和，且这些初等因子是系数取自 **F** 的多项式.

例如，考察矩阵

$$A_1=[1],\quad A_2=\begin{bmatrix}0&-4\\1&4\end{bmatrix},\quad A_3=\begin{bmatrix}0&-9\\1&0\end{bmatrix},\quad A_4=\begin{bmatrix}4&0\\0&4\end{bmatrix},$$

且设 $A=A_1\oplus A_2\oplus A_3\oplus A_3\oplus A_4\in M_9$. 那么 A 在 **R** 上的有理标准形是 $A=A_1\oplus A_2\oplus A_3\oplus A_3\oplus[4]\oplus[4]$，且初等因子为 $x-1$，$(x-2)^2$，x^2+9，x^2+9，$x-4$，$x-4$. 在 **C** 上，A 的有理标准形是 $A_1\oplus A_2\oplus[3i]\oplus[-3i]\oplus[-3i]\oplus[4]\oplus[4]$，且初等因子为 $x-1$，$(x-1)^2$，$x-3i$，$x-3i$，$x+3i$，$x+3i$，$x-4$，$x-4$. A 在 **R** 上的有理形是($A_1\oplus A_2\oplus A_3\oplus[4]$的友矩阵)$\oplus$($A_3\oplus[4]$的友矩阵)，且不变因式是

$$f_1(t)=(t-1)(t-2)^2(t^2+9)(t-4)\ \text{和}\ f_2(t)=(t^2+9)(t-4).$$

在 **C** 上，A 的有理形是($A_1\oplus A_2\oplus[3i]\oplus[-3i]\oplus[4]$的友矩阵)$\oplus$($[3i]\oplus[-3i]\oplus[4]$)的友矩阵). 应注意的是，不论把 A 看成是 $M_n(\mathbf{R})$中的，还是 $M_n(\mathbf{C})$中的矩阵，这两个直加矩阵以及不变因式都相同，这对于有理标准形及其初等因子是不成立的. 见本节末的习题 2 和 3.

在本节的其他地方，基本上不采用实 Jordan 形、有理形、有理标准形、不变因式或初等因子，在这里讨论它们，仅是因为它们在历史上的重要作用，同时还因为在不同于 **C** 的域上作矩阵分析时，它们是必不可少的.

155

还有许多其他有用的标准形和矩阵分解：

(a) 极分解：每个 $A\in M_n$ 可以写成 $A=PU$，其中 $A\in M_n$ 是与 A 有相同秩的半正定矩阵，[136] 而 $U\in M_n$ 是酉矩阵．见(7.3.3)．每个非奇异矩阵 $A\in M_n$ 也可以写成 $A=GQ$，其中，$G\in M_n$ 是(复)对称矩阵($G=G^T$)，且 $Q\in M_n$ 是复正交矩阵($QQ^T=I$)．

(b) 奇异值分解：每个 $A\in M_n$ 可以写成 $A=V\Sigma W^*$，其中，V，$W\in M_n$ 是酉矩阵，而 $\Sigma\in M_n$ 是具有非负主对角元的对角矩阵，且 Σ 的秩与 A 的秩相同．见(7.3.5)．

(c) 三角分解：每个 $A\in M_n$ 可以写成 $A=URU^*$，其中，$U\in M_n$ 是酉矩阵，而 $R\in M_n$ 是上三角矩阵．每个实矩阵 $A\in M_n(\mathbf{R})$ 可以写成 $A=QRQ^T$，其中 $Q\in M_n(\mathbf{R})$ 是正交矩阵，而 $R\in M_n(\mathbf{R})$ 是具有一个特殊结构的上 Hessenberg 矩阵．见(2.3.5)．

(d) 每个 Hermite 矩阵 $A\in M_n$ 可以写成 $A=SI(A)S^*$，其中，$S\in M_n$ 是非奇异矩阵，而 $I(A)\in M_n$ 是以 $+1$，-1，或 0 为主对角元的对角矩阵．在 $I(A)$ 中，元素 $+1(-1)$ 的个数与 A 的正(负)特征值的个数是相同的；0 元的个数等于 $n-\operatorname{rank} A$．见(4.5.8)．

(e) 每个正规矩阵 $A\in M_n$ 可以写成 $A=U\Lambda U^*$，其中 $U\in M_n$ 是酉矩阵，而 $\Lambda\in M_n$ 是其主对角元为 A 的各特征值的对角矩阵．每个实正矩阵 $A\in M_n(\mathbf{R})$ 可以写成 $A=QDQ^T$，其中 $Q\in M_n(\mathbf{R})$ 是正交矩阵，而 $D\in M_n(\mathbf{R})$ 是具有一个特殊结构的分块对角矩阵．见(2.5.8)．

(f) 每个使得 $A=A^T$ 的矩阵 $A\in M_n$ 可以写成 $A=SK(A)S^T$，其中，$S\in M_n$ 是非奇异矩阵，而 $K(A)\in M_n$ 是主对角元为 +1 或 0，且其秩等于 A 的秩的对角矩阵，见(4.5.12)．

(g) 每个使得 $A=A^T$ 的矩阵 A 可以写成 $A=U\Sigma U$，其中，$U\Sigma M_n$ 是酉矩阵，而 Σ 是具有非负主对角元的对角矩阵，Σ 的秩等于 A 的秩．见(4.4.4)．

(h) 每个酉矩阵 $U\in M_n$ 可以写成 $U=Qe^{iE}$，而每个(复)正交矩阵 $P\in M_n$ 可以写成 $P=Qe^{iF}$，其中 Q，E，$F\in M_n(\mathbf{R})$，Q 是实正交矩阵($QQ^T=I$)，E 是实对称矩阵($E=E^T$)，且 F 是实斜对称矩阵($F=-F^T$)．

(i) 每个矩阵 $A\in M_n$ 可以写成 $A=SU\Sigma U^TS^{-1}$，其中 S 是非奇异矩阵，U 是酉矩阵，而 Σ 是具有非负主对角元的对角矩阵．见(4.4.10)．

习题

[157] 1. 对于

$$\begin{bmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \quad 和 \quad \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

在 **R** 和 **C** 上计算它们的极小多项式，特征多项式，不变因式，初等因子，有理形和有理标准形．

2. 设 $A\in M_n(\mathbf{R})$．假定 $q(t)$ 是 A 在 **R** 上的极小多项式，而 $f(t)$ 是 A 在 **C** 上的极小多项式．为什么 $f(t)$ 的次数 $\leqslant q(t)$ 的次数？为什么 $f(t)$ 一定除尽 $q(t)$？假定 $f(t)=p_1(t)+ip_2(t)$，其中 $p_1(t)$ 和 $p_2(t)$ 有实系数．试证 $f(t)=q(t)$．为什么 $p_1(A)=p_2(A)=0$？

3. 试用定理(3.4.7)证明，如果 A，$B\in M_n(\mathbf{F})$，且 **F** 是 **C** 的一个子域(例如 **F**=**R** 或 **Q**)，那么 A 与 B 在 **F** 上相似，当且仅当它们在 **C** 上相似．**提示**：试证 A 在 **F** 上的有理形与 A 在 **C**

上的有理形是相同的，且对于 B 也有类似的结果．这一结论是如何推广习题 2 的?

4．设 $A\in M_n(\mathbf{R})$，且假定 $A^2=-I$，证明，n 必须是偶数，且有一个非奇异实矩阵 $S\in M_n$，使得

$$S^{-1}AS=\begin{bmatrix}0 & -I\\ I & 0\end{bmatrix},$$

其中每个单位矩阵 $I\in M_{n/2}$.

进一步阅读 本节所论述的有理标准形是十分经典的，其更详细的论述可以在[HKu]中找到．在一段时间里，实 Jordan 标准形也为从事矩阵理论研究的人所熟知，但有关它的文献不多见．例如，关于实 Jordan 形的论述可在[Kow]中找到．关于其元素是有理数或整数的矩阵的标准形的讨论可见[New]．关于特殊标准形的更为详细的讨论可见[Gan]，vol. 2；[Gant]和[HJ].

3.5 三角分解

如果线性方程组 $Ax=b$ 有一个非奇异三角(0.9.3)系数矩阵 $A\in M_n$，那么计算它的唯一解 x 是相当容易的．例如，如果 A 是上三角矩阵

$$A=\begin{bmatrix}a_{11} & \cdots & a_{1n}\\ & a_{22} & \\ & \ddots & \vdots\\ 0 & & a_{nn}\end{bmatrix},$$

158

那么 $\det A=a_{11}a_{22}\cdots a_{nn}\neq 0$，然后利用后向替换：$a_{nn}x_n=b_n$ 确定 x_n；$a_{n-1,n-1}x_{n-1}+a_{n-1,n}x_n=b_{n-1}$是含一个未知数的方程，它确定 x_{n-1}；一般，方程组

$$\sum_{j=i}^{n}a_{ij}x_j=b_i,\quad i=n,n-1,n-2,\cdots,2,1$$

中的每一个是含一个未知数的方程(只要 x_{i+1}，…，x_n 已被确定)，它确定 x_i.

练习 设 $A\in M_n$ 是非奇异上三角矩阵，如果采用后向替换，试算出解 $Ax=b$ 所必需的纯量乘法和除法运算的次数.

练习 如果 $A\in M_n$ 是非奇异下三角矩阵，写出作为 $Ax=b$ 的一个解法的前向替换.

设 $A\in M_n$ 是非奇异矩阵，但不是三角矩阵，应指出的是，如果 A 是用形如

$$A=LU$$

的分解形式给出的，其中，L 是下三角矩阵，U 是上三角矩阵，那么解出 $Ax=b$ 可以说是很简便的.

练习 如果 $A=LU$ 如上所述且非奇异，证明，L 和 U 都必须是非奇异的，因而必须有非零对角元.

为了解 $Ax=b$，可以先用前向替换解

$$Ly=b,$$

然后用后向替换解

$$Ux = y,$$

且所需计算量只是系数矩阵为简单的三角矩阵时的两倍. 因此, 如果得到分解 LU 所需计算量不太大, 那么这种分解对解线性方程组是有用的. 在这里叙述它们也是适宜的, 因为它们的形式特殊, 而把一个矩阵分解成这种形式可以不用特征值, 而是利用线性方程组.

3.5.1 引理 假定 $A\in M_n$ 可以写成

159

$$A = LU,$$

其中 $L\in M_n$ 是下三角矩阵, 而 $U\in M_n$ 是上三角矩阵. 对任一分块形式

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},\quad L=\begin{bmatrix}L_{11} & 0\\ L_{21} & L_{22}\end{bmatrix},\quad U=\begin{bmatrix}U_{11} & U_{12}\\ 0 & U_{22}\end{bmatrix},$$

其中 A_{11}, L_{11}, $U_{11}\in M_k$, $k\leqslant n$, 我们有

$$L_{11}U_{11} = A_{11},$$

$$L_{11}U_{12} = A_{12},\quad L_{21}U_{11} = A_{21},$$

且

$$L_{21}U_{12} + L_{22}U_{22} = A_{22}.$$

特别是, L 和 U 的左上角子块一定构成 A 的相应子块的一个同型的分解.

练习 用分块乘法运算验证(3.5.1).

3.5.2 定理 假定 $A\in M_n$, 且 $\operatorname{rank} A=k$. 如果

$$\det A(\{1,\cdots,j\}) \neq 0,\quad j = 1,\cdots,k,$$

那么 A 可以分解成

$$A = LU,$$

其中, $L\in M_n$ 是下三角矩阵, $U\in M_n$ 是上三角矩阵. 并且可以适当选择分解使得 L 或 U 是非奇异的; L 和 U 都可以选取非奇异矩阵, 当且仅当 $k=n$, 即当且仅当 A 是非奇异的.

证明: 首先证明, 在关于前主子式的假设条件下, $A(\{1,\cdots,k\})$ 可以分解成 $L(\{1,\cdots,k\})U(\{1,\cdots,k\})$, 且它们都是非奇异的. 能够逐个地求出 L 和 U 的各相关元素. 设 $L=[l_{ij}]$ 且 $U=[u_{ij}]$. 令 $u_{11}=1$ 且令 $l_{i1}=a_{i1}$, $i=1,\cdots,k$. 求出

$$u_{1j} = \frac{a_{1j}}{l_{11}},\quad j = 2,\cdots,k.$$

继续做下去. 令 $u_{22}=1$, 且令 $l_{i2}=a_{i2}-l_{i1}u_{12}$, $i=2,\cdots,k$. 求出

$$u_{2j} = \frac{a_{2j} - l_{21}u_{1j}}{l_{22}},\quad j = 3,\cdots,k.$$

再继续下去, 依次设 U 的各对角元为 1, 然后求出 $L(\{1,\cdots,k\})$ 的下一列和 $U(\{1,\cdots,k\})$ 的下一行. 每次都有一个含一个未知数的方程需要求解. 因为每个 l_{ii} 不为零[因为根据(3.5.1), $\det L(\{1,\cdots,i\})\times\det U(\{1,\cdots,i\})=\det A(\{1,\cdots,i\})$], 所以这个方程将是可解的. 这就完成了 $A(\{1,\cdots,k\})$ 的分解.

160

将 A 如(3.5.1)中那样分块. 因为 $\operatorname{rank} A=k=\operatorname{rank} A_{11}$, 由此可知$[A_{21}\, A_{22}]$的各行是$[A_{11}\, A_{12}]$的诸行的唯一线性组合, 即对每个唯一确定的 $B\in M_{n-k,k}$, 有

$$A_{21} = BA_{11} \quad 和 \quad A_{22} = BA_{12}.$$

现在将欲求的 L 和 U 同样如(3.5.1)中那样分块，注意到非奇异的 L_{11} 和 U_{11} 已被确定．于是可以用(3.5.1)求出

$$U_{12} = L_{11}^{-1}A_{12} \quad 和 \quad L_{21} = A_{21}U_{11}^{-1}.$$

然后求出

$$A_{22} = L_{21}U_{12} + L_{22}U_{22} = A_{21}U_{11}^{-1}L_{11}^{-1}A_{12} + L_{22}U_{22} = BA_{11}A_{11}^{-1}A_{12} + L_{22}U_{22}$$
$$= A_{22} + L_{22}U_{22}.$$

为了完成分解，必需而且只需

$$L_{22}U_{22} = 0.$$

例如，可以选取 L_{22}(相应地，U_{22})为 M_{n-k} 中的任一非奇异下(相应地，上)三角矩阵，希望选取 U_{22}(相应地，L_{22})为 0．因为 L_{11} 和 U_{12} 是非奇异的，L 或 U 可以选为非奇异的．如果 $k=n$，$L=L_{11}$ 和 $U=U_{11}$ 就是可非奇异的；如果 $k<n$，因为 A 是奇异的，L 和 U 不可能都是非奇异的．这就完成了证明． □

3.5.3 例 不是每个矩阵都有一个 LU 分解．考虑

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

如果 A 可以写成

$$A = LU = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix}\begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix},$$

$l_{11}u_{11}=0$ 将要求 L 或 U 是奇异的，但是 $LU=A$ 是非奇异的．

练习 证明，一个非奇异矩阵的左上角有一个 $k\times k$ 奇异主子阵，它就不可能有 LU 分解． 161

3.5.4 例 $A\in M_n$ 可以有 LU 分解而不满足(3.5.2)的主子式条件．例如

$$\begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

有秩 1，但是它的 1，1 元是 0.

练习 (3.5.4)中的 LU 的分解是不唯一的，即使要求 U 的各对角元都为 1．试写出 $\begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix}$ 的其他几种分解．

现在已经很清楚了，一个已知矩阵的 LU 分解是不唯一的，它可能存在，或可能不存在．不过，这许多麻烦，或者是由 A 的奇异性，或者是由 A 的前主子阵的奇异性引起的．然而可以采用(3.5.1)和(3.5.2)的方法，对非奇异的情形给出一个完美的描述，并且可以利用规范化使分解是唯一的(标准的)．

3.5.5 推论 假定 $A\in M_n$ 是非奇异的．那么，A 可以写成

$$A = LU,$$

且使 $L\in M_n$ 是下三角矩阵和 $U\in M_n$ 为上三角矩阵，当且仅当

$$\det A(\{1,\cdots,j\}) \neq 0, \quad j = 1,\cdots,n.$$

此外，L 和 U 是非奇异的，而分解实质上是唯一的. 矩阵 A 可以写成

$$A = L'DU',$$

其中，L'(相应地，U')$\in M_n$ 是所有对角元等于1的下(相应地，上)三角矩阵，而 D 是由

$$\det D(\{1,\cdots,j\}) = \det A(\{1,\cdots,j\}), \quad j = 1,\cdots,n$$

所确定的非奇异对角矩阵. 因子 L'，U'和 D 被 A 唯一确定.

练习 利用(3.5.1)，(3.5.2)和前一个练习，详细证明(3.5.5).

再回到线性方程组

162

$$Ax = b$$

的解法，假定 $A\in M_n$ 不能分解成 LU，但能分解成 PLU，其中，$P\in M_n$ 是置换矩阵(0.9.5)，而 L 和 U 如前，是下三角矩阵和上三角矩阵. 这相当于在分解之前重排各个方程. 在这种情形，$Ax=b$ 的解法仍然是相当简单的，只要作替换

$$Ly = P^Tb \quad 和 \quad Ux = y$$

就是了. 应当知道，任一非奇异的 $A\in M_n$ 都可以这样分解. 而且任一 $A\in M_n$ 可以分解成 $PLUQ$，其中 $Q\in M_n$ 也是一个置换矩阵.

3.5.6 引理 设 $A\in M_k$ 是非奇异矩阵. 那么，存在一个置换矩阵 $P\in M_k$，使得

$$\det(P^TA)(\{1,\cdots,j\}) \neq 0, \quad j = 1,\cdots,k.$$

注意，P^TA 正好是 A 的各解的一个重排.

证明： 证明是对 k 作归纳法. 如果 $k=1$ 或 2，经过验证，结论显然成立；假定结论一直到 $k-1$ 都是正确的. 考虑非奇异矩阵 $A\in M_k$，并且去掉它的最后一列. 余下的 $k-1$ 列是线性无关的，因而含有 $k-1$ 个线性无关解. 把这些解调换到前 $k-1$ 个解的位置，然后对位于上方的非奇异的$(k-1)\times(k-1)$子矩阵应用归纳假设. 这就确定了一个欲求的整个置换矩阵. 因为 P^TA 是非奇异的，所以证明完毕. □

3.5.7 定理 设 $A\in M_n$，则存在置换矩阵 P，$Q\in M_n$，下三角矩阵 $L\in M_n$ 和上三角矩阵 $U\in M_n$，使得

$$A = PLUQ.$$

如果 A 是非奇异矩阵，可以取 $Q=I$，且 A 可以写成

$$A = PLU.$$

证明： 如果 rank $A=k$ 则 A 有一个 $k\times k$ 非奇异子矩阵(0.4.4d)，它可以通过互换 A 的各行和各列调换到左上角的位置. 现在把(3.5.6)应用于左上角，再应用(3.5.2)便得到第一型式的分解. 如果 A 是非奇异的，(3.5.6)表明，为应用(3.5.2)，右边的置换矩阵是不必要的，这说明第二个分解是正确的，证毕. □

163

习题

1. 本节所提出的理论是就 LU 分解而论的，其中 L 是下三角矩阵，U 是上三角矩阵. 试说明，可以平行地提出 UL 分解的理论，但这两个因子一般与 LU 分解的因子不相同.

2. 从(2.6)节习题3可知，任一 $A\in M_n$ 的 QR 分解可以有效地经 $n-1$ 次 Householder 交

换得到. 这里，Q 是酉矩阵，而 R 是上三角矩阵. 如果 A 分解成 QR 形，试描述可以怎样解出 $Ax=b$.

3. 证明，$A\in M_n$ 可以写成

$$A = LP_0U,$$

其中 $L\in M_n$ 是非奇异下三角矩阵，$U\in M_n$ 是非奇异上三角矩阵，而 P_0 是一个次置换矩阵[当 rank $A<n$ 时，置换矩阵中的一些元素 1 换成元素 0，使其元素 1 的个数与 rank A 相同]. **提示**：进行初等行变换和初等列变换.

4. 如果 $A\in M_n$ 的所有前主子式都非零，试描述如何运用第三种初等行变换将 A 的对角线下方的各个元素都化成零，从而使 A 有 LU 分解.

5. (Lanczos 三对角化算法.) 设 $A\in M_n$ 和 $x\in \mathbf{C}^n$ 都是给定的. 定义 $X=[x\ Ax\ A^2x\cdots A^{n-1}x]$. 称 X 的诸列构成一个 krylov 序列. 假定 X 是非奇异的. (a)证明 $X^{-1}AX$ 是 A 的特征多项式的友矩阵(3.3.12). (b)若 $R\in M_n$ 是任意一个给定的非奇异上三角矩阵且 $S\equiv XR$，证明 $S^{-1}AS$ 呈上 Hessenberg 形式. (c)设 $y\in \mathbf{C}^n$ 且定义 $Y=[y\ A^*y\ (A^*)^2y\cdots(A^*)^{n-1}y]$. 假定 Y 是非奇异的且 Y^*X 可以写成 LDU，其中，L 是下三角矩阵，U 是上三角矩阵，且都是非奇异的，而 D 是对角矩阵且是非奇异的. 证明，存在非奇异上三角矩阵 R 和 T，使得 $(XR)^{-1}=T^*Y^*$ 且使 T^*Y^*AXR 是相似于 A 的三对角矩阵. (d)若 $A\in M_n$ 是 Hermite 矩阵，利用上述思路确定一个算法来得到一个相似于 A 的三对角 Hermite 矩阵.

6. 两个矩阵 A，$B\in M_{m,n}$ 称为等价，指的是存在非奇异矩阵 $S\in M_m$ 和 $T\in M_n$，使得

$$B = SAT.$$

(a) 证明这个等价概念是 $M_{m,n}$ 上的一个等价关系. (b)证明，每个矩阵 $A\in M_{m,n}$ 等价于一个形如 $\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\in M_{m,n}$ 的矩阵，$I\in M_k$，$k\leqslant \min\{m, n\}$. **提示**：利用初等行变换得到行简化梯形， 164 然后对所得到的矩阵采用初等列变换. (c)证明，$M_{m,n}$ 中的两个矩阵等价，当且仅当它们有相同的秩. (d)假定 $A\in M_{m,n}$ 等价于(b)中所述的特殊形式，$S\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}T=A$. 试用等价性阐述关于线性方程组 $Ax=b$ 的解理论.

进一步阅读　上述习题 5 是从[Ste]改编过来的，在[Ste]中还可以找到关于 LU 分解的数值应用的其他资料. 165

第 4 章　Hermite 矩阵和对称矩阵

4.0　导引

4.0.1 例　如果 $f: D \to \mathbf{R}$ 是某个域 $D \subset \mathbf{R}^n$ 上的二次连续可微函数，实矩阵

$$H(x) = [h_{ij}(x)] \equiv \left[\frac{\partial^2 f(x)}{\partial x_i \, \partial x_j}\right] \in M_n$$

称为 f 的 Hessian 矩阵．它是 x 的函数，因为可以用它来确定一个临界点是否为相对极大值点或极小值点，所以它在最优化理论中起着重要的作用[（见 7.0）]．

目前，使我们特别感兴趣的 $H = H(x)$ 性质来源于混合偏导数相等的重要事实；这就是对所有 i，$j = 1, 2, \cdots, n$ 有

$$\frac{\partial^2 f}{\partial x_i \, \partial x_j} = \frac{\partial^2 f}{\partial x_j \, \partial x_i}.$$

用 Hessian 矩阵 $H = [h_{ij}]$ 来表示，这意味着，对所有 i，$j = 1, 2, \cdots, n$，有 $h_{ij} = h_{ji}$；即 $H = H^T$．设矩阵 $A \in M_n$，如果有 $A = A^T$，就称它是**对称矩阵**．因此，一个二次连续可微的实值函数的 Hessian 矩阵总是实对称矩阵．

4.0.2 例　作为第二个例子，设 $A = [a_{ij}] \in M_n$ 是某个具有实或复元素的矩阵，考虑由 A 产生的 $\mathbf{R}^n$ 或 $\mathbf{C}^n$ 上的二次型：

167

$$\begin{aligned} Q(x) = x^T A x &= \sum_{i,j=1}^{n} a_{ij} x_i x_j \\ &= \sum_{i,j=1}^{n} \frac{1}{2}(a_{ij} + a_{ji}) x_i x_j \\ &= x^T \left[\frac{1}{2}(A + A^T)\right] x. \end{aligned}$$

于是，A 和 $\frac{1}{2}(A + A^T)$ 引出同一个二次型，且后一个矩阵是对称的．因此，为了研究实的或复的二次型，只要研究由对称矩阵产生的那些二次型就可以了．例如，在物理中，作为物体的惯性表达式，自然会遇到实二次型．

4.0.3 例　作为第三个例子，考虑用

$$Lf(x) = \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \, \partial x_j} \tag{4.0.4}$$

定义的二阶线性偏微分算子 L．假定系数 $a_{ij}(x)$ 和函数 $f(x)$ 都是定义在同一个区域 $D \subset \mathbf{R}^n$ 上的，且 f 在 D 上应当是二次可微的．可以用一种自然的方式把算子 L 与一个矩阵联系起来，矩阵 $A = [a_{ij}(x)]$ 不一定是对称的，但是，因为 f 的混合偏导数相等，因而有

$$Lf = \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial^2 f}{\partial x_i \, \partial x_j} = \sum_{i,j=1}^{n} \frac{1}{2}\left[a_{ij}(x) \frac{\partial^2 f}{\partial x_i \, \partial x_j} + a_{ji}(x) \frac{\partial^2 f}{\partial x_j \, \partial x_i}\right]$$

$$= \sum_{i,j=1}^{n} \frac{1}{2}[a_{ij}(x) + a_{ji}(x)] \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

因此，对称矩阵 $\frac{1}{2}(A+A^T)$ 与矩阵 A 产生同一个算子 L，并且为了研究形如(4.0.4)的实的或复的线性偏微分算子，只要考虑对称系数矩阵就够了.

4.0.5 例 考虑无向图 Γ：即 Γ 由集合 N 和集合 E 组成，其中 N 是"结点"的集合 $\{P_1, P_2, \cdots, P_n\}$，$E$ 是无序结点偶(称为"边")的集合 $E=\{\{P_{i_1}, P_{j_1}\}, \{P_{i_2}, P_{j_2}\}, \cdots\}$. 图 Γ 可以简洁地用它的所谓邻接矩阵 $A=[a_{ij}]$ 来描述，这里

$$a_{ij} = \begin{cases} 1, & \text{如果}\{P_i, P_j\} \in E, \\ 0, & \text{否则}. \end{cases}$$

168 因为 Γ 是无向图，所以 A 是实对称矩阵：即 $A^T=A$.

4.0.6 例 设 $A=[a_{ij}]\in M_n$ 是实矩阵，考虑实双线性型

$$Q(x,y) = y^TAx = \sum_{i,j=1}^{n} a_{ij}y_ix_j, \quad x,y \in \mathbf{R}^n. \tag{4.0.7}$$

当 $A=I$ 时，它简化为普通内积. 如果希望对所有 x，y，有 $Q(x, y)=Q(y, x)$，那么它必需而且只需对所有 i，$j=1$，…，n，有 $a_{ij}=a_{ji}$. 为了证明这一点，只需注意到，如果 $x=e_j$ 和 $y=e_i$，那么 $Q(e_j, e_i)=a_{ij}$ 且 $Q(e_i, e_j)=a_{ji}$. 因此，实对称双线性型自然与实对称矩阵相对应.

现在设 $A=[a_{ij}]\in M_n$ 是实的或复的矩阵，考虑复形式

$$H(x,y) = y^*Ax = \sum_{i,j=1}^{n} a_{ij}\bar{y}_ix_j, \quad x,y \in \mathbf{C}^n. \tag{4.0.8}$$

同(4.0.7)一样，当 $A=I$ 时，它简化为普通内积. 这种形式不再是双线性的，但是它对第一变元是线性的，而对第二变元是"共轭线性"的($H(ax, by)=a\bar{b}H(x, y)$)，这正好与复 Euclid 内积相同. 有人称这种形式为半双线性的. 如果希望像内积一样有 $H(x, y)=\overline{H(y, x)}$，那么，与前述情形相同的论证说明，它必需而且只需有 $a_{ij}=\bar{a}_{ji}$；即 $A=\bar{A}^T\equiv A^*$. 要注意的是，如果 A 是实矩阵，那么 $A^*=A^T$.

使 $A=A^*$ 的 $A\in M_n$ 的矩阵类在许多方面是实对称矩阵类到 $M_n(\mathbf{C})$ 的自然推广. 这样的矩阵称为 Hermite 矩阵；要注意的是，实 Hermite 矩阵就是实对称矩阵. 非实的复对称矩阵类没有实对称矩阵类那么多重要性质. 在本章，我们将研究复 Hermite 矩阵和复对称矩阵，并且要特别指出在实对称情形成立的那些性质.

4.1 Hermite 矩阵的定义、性质和特征

4.1.1 定义 矩阵 $A=[a_{ij}]\in M_n$ 称为 Hermite 矩阵，是指 $A=A^*$，其中 $A^*\equiv\bar{A}^T=[\bar{a}_{ji}]$. 如果 $A=-A^*$，则称之为斜 Hermite 矩阵.

关于 A，$B\in M_n$ 的一些论断：

1. 对所有 $A\in M_n$，$A+A^*$，AA^* 和 A^*A 都是 Hermite 矩阵.
2. 如果 A 是 Hermite 矩阵，那么，对所有 $k=1$，2，3，…，A^k 是 Hermite 矩阵. 如果 A

还是非奇异的，那么，A^{-1} 也是 Hermite 矩阵. [169]

3. 如果 A，B 是 Hermite 矩阵，那么，对所有实纯量 a，b，$aA+bB$ 是 Hermite 矩阵.

4. 对所有 $A\in M_n$，$A-A^*$ 是斜 Hermite 矩阵.

5. 如果 A，B 是斜 Hermite 矩阵，那么，对所有实纯量 a，b，$aA+bB$ 是斜 Hermite 矩阵.

6. 如果 A 是 Hermite 矩阵，那么 iA 是斜 Hermite.

7. 如果 A 是斜 Hermite 矩阵，那么 iA 是 Hermite 矩阵.

8. 任意 $A\in M_n$ 可写成

$$A=\frac{1}{2}(A+A^*)+\frac{1}{2}(A-A^*)=H(A)+S(A),$$

其中 $H(A)=\frac{1}{2}(A+A^*)$ 是 A 的 Hermite 部分，而 $S(A)=\frac{1}{2}(A-A^*)$ 是 A 的斜 Hermite 部分.

9. 如果 A 是 Hermite 矩阵，那么 A 的主对角元都是实数，为了给出 A 的 n^2 个元素，我们可以自由地给定任意 n 个实数(对于主对角元)和任意 $\frac{1}{2}n(n-1)$ 个复数(对于非对角元).

4.1.2 定理 每个 $A\in M_n$ 可以唯一地写成 $A=S+iT$，其中 S 和 T 都是 Hermite 矩阵. 它也可以唯一地写成 $A=B+C$，其中 B 是 Hermite 矩阵，而 C 是斜 Hermite 矩阵.

证明： 把 A 写成 $A=\frac{1}{2}(A+A^*)+i[(-i/2)(A-A^*)]$，并且注意到 $S=\frac{1}{2}(A+A^*)$ 和 $T=(-i/2)(A-A^*)$ 都是 Hermite 矩阵. 关于唯一性论断，我们知道，如果 $A=E+iF$，其中 E 和 F 都是 Hermite 矩阵，那么

$$2S=A+A^*=(E+iF)+(E+iF)^*=E+iF+E^*-iF^*=2E,$$

因而 $E=S$. 类似地可以证明，$F=T$. 关于表示式 $A=B+C$ 的论断也可用同样的方式来证明. □

上述论断使我们联想到，如果把 M_n 比作复数，那么 Hermite 矩阵就可以比作实数. $\mathbf{C}$ 中的复共轭运算类似于 M_n 上的(伴随)运算*. 一个实数是使 $z=\bar{z}$ 的复数 z；一个 Hermite 矩阵是使 $A=A^*$ 的矩阵 $A\in M_n$. 正如每个复数 z 可以写成 $z=s+it$ 一样(其中 a，$t\in\mathbf{R}$)，每个复矩阵 A 也可以唯一地写成 $A=S+iT$，其中 S 和 T 是 Hermite 矩阵，还有另外一些性质进一步说明这种类似性.

4.1.3 定理 设 $A\in M_n$ 是 Hermite 矩阵，那么

(a) 对所有 $x\in\mathbf{C}^n$，x^*Ax 是实数；

(b) A 的所有特征值都是实数；

(c) 对所有 $S\in M_n$，S^*AS 是 Hermite 矩阵. [170]

证明： 计算 $\overline{(x^*Ax)}=(x^*Ax)^*=x^*A^*x=x^*Ax$，于是 x^*Ax 等于它的复共轭，因而它是实数. 如果 $Ax=\lambda x$ 且 $x^*x=1$，那么，根据(a)，$\lambda=\lambda x^*x=x^*\lambda x=x^*Ax$ 是实数. 最后，$(S^*AS)^*=S^*A^*S=S^*AS$，所以 S^*AS 总是 Hermite 矩阵. □

练习 当 $n=1$ 时，Hermite 矩阵 $A\in M_n$ 的上述每个性质指的是什么？

(4.1.3)的每一个性质实际上(几乎都)是 Hermite 矩阵的一个特征.

4.1.4 定理 设 $A=[a_{ij}]\in M_n$ 是给定的，那么，A 是 Hermite 矩阵，当且仅当下述条件至少有一个成立：

(a) 对所有 $x\in \mathbf{C}^n$，x^*Ax 是实数；

(b) A 是正规矩阵，且 A 的所有特征值都是实数；

(c) 对所有 $S\in M_n$，S^*AS 是 Hermite 矩阵.

证明：这只要证明每个条件的充分性就可以了. 如果对所有 $x\in \mathbf{C}^n$，x^*Ax 是实数，那么，对所有 x，$y\in \mathbf{C}^n$，$(x+y)^*A(x+y)=(x^*Ax+y^*Ay)+(x^*Ay+y^*Ax)$ 是实数. 因为根据假设条件，x^*Ax+y^*Ay 是实数，得出，对所有 x，$y\in \mathbf{C}^n$，x^*Ay+y^*Ax 是实数. 如果选取 $x=e_k$ 和 $y=e_j$，这就是说，$a_{kj}+a_{jk}$ 是实数，因而 $\mathrm{Im}\ a_{kj}=-\mathrm{Im}\ a_{jk}$. 如果选取 $x=e_k$ 和 $y=ie_j$，这就是说，$-ia_{kj}+ia_{jk}$ 是实数，因而 $\mathrm{Re}\ a_{kj}=\mathrm{Re}\ a_{jk}$. 这两个恒等式结合起来就等价于有 $a_{kj}=\bar{a}_{jk}$，又因为 j，k 是任意的，因此得出 $A=A^*$.

如果 A 是正规矩阵，它就是可酉对角化的，所以 $A=U\Lambda U^*$，其中，$\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$ 是由 A 的各特征值构成的对角矩阵. 一般地，有 $A^*=U\bar{\Lambda}U^*$，但是，如果 Λ 是实矩阵，就有 $A^*=U\Lambda U^*=A$. 最后一个条件推出 A 是 Hermite 矩阵，这只需取 $S=I$. □

因为 Hermite 矩阵显然是正规矩阵($AA^*=A^2=A^*A$)，第 2 章中有关正规矩阵的所有结果都适用于 Hermite 矩阵. 例如，相应于不同特征值的特征向量是正交的. 存在正交特征向量的一个完备集；Hermite 矩阵是可以酉对角化的；等等.

作为参考，我们特别叙述下述重要结果.

4.1.5 定理(Hermite 矩阵的谱定理) 设 $A\in M_n$ 是给定的，那么，A 是 Hermite 矩阵，当且仅当存在一个酉矩阵 $U\in M_n$ 和一个实对角矩阵 $\Lambda\in M_n$，使得 $A=U\Lambda U^*$. 此外，A 是实 Hermite 矩阵(即实对称的)，当且仅当存在一个实正交矩阵 $P\in M_n$ 和一个实对角矩阵 $\Lambda\in M_n$，使得 $A=P\Lambda P^T$.

[171]

虽然 Hermite 矩阵的实线性组合总是 Hermite 矩阵，但它们的复线性组合就不一定是 Hermite 矩阵. 例如，如果 A 是 Hermite 矩阵，那么，只有当 $A=0$ 时 iA 才是 Hermite 矩阵. 另外，如果 A 和 B 是 Hermite 矩阵，那么 $(AB)^*=B^*A^*=BA$，因此，AB 是 Hermite 矩阵，当且仅当 A 与 B 可交换.

关于交换的 Hermite 矩阵的最著名的结果之一(因为在量子力学中它是到算子的重要推广)是定理(2.5.5)的下述特殊情形.

4.1.6 定理 设 $\mathcal{F}$ 是给定的 Hermite 矩阵族. 对所有 $A\in\mathcal{F}$，存在酉矩阵 U，使得 UAU^* 是对角阵，当且仅当对所有 A，$B\in\mathcal{F}$ 有 $AB=BA$.

Hermite 矩阵 A 有 A 等于 A^* 的性质，推广 Hermite 矩阵概念的一种方式是考察 A 相似于 A^* 的矩阵类. 下述定理用几种方式刻划了这一类矩阵，其中第一个是说，这样的矩阵必定相似于(但不一定酉相似于)一个实的(但不一定是对角的)矩阵.

4.1.7 定理 设 $A\in M_n$ 是给定的，下述诸命题等价：

(a) A 相似于矩阵 $B\in M_n(\mathbf{R})$；

(b) A 相似于 A^*；

(c) A 经 Hermite 相似变换相似于 A^*；

(d) $A=HK$，其中 H，$K\in M_n$ 都是 Hermite 矩阵，且至少有一个是非奇异的；

(e) $A=HK$，其中 H，$K\in M_n$ 是 Hermite 矩阵.

证明：首先要指出的是(a)和(b)等价：如果(a)成立，则 $S^{-1}AS=B=T^{-1}B^TT=T^{-1}B^*T=T^{-1}S^*A^*(S^{-1})^*T$，这就是说，$A^*=(ST^{-1}S^*)^{-1}A(ST^{-1}S^*)$，或者说(b)成立. 如果(b)成立，那么 A 和 A^* 有相同的 Jordan 标准形. 因为对任意矩阵 A，A 与 A^T 相似，这说明，如果 J 是 A 的 Jordan 矩阵，那么 J 必须相似于 $\bar{J}$. 因此，对 J 中的每个 Jordan 块 $J_k(\lambda)$，在 $\bar{J}$ 中有一个相应的(相同阶数的)Jordan 块 $J_k(\bar{\lambda})$. 如果 λ 是实数，就没什么可说的了，如果 λ 不是实数，这就是说，相应于每个非实特征值的诸 Jordan 块和它的共轭必须成对出现. 利用推导(3.4.5)的论证，得出 J 必须相似于形如(3.4.4)的实矩阵的一个直和，因而(a)成立. 172

为了证明(b)蕴涵(c)，假定 $S^{-1}AS=A^*$，且注意到，如果对任意非零 $\alpha=re^{i\theta}\in\mathbf{C}$，$T=\alpha S$，那么 $T^{-1}AT=A^*$. 因而 $AT=TA^*$，或等价地 $AT^*=T^*A^*$. 相加这两个恒等式得恒等式 $A(T+T^*)=(T+T^*)A^*$，而如果 $T+T^*$ 是非奇异矩阵，这便说明 A 可经 Hermite 矩阵 $T+T^*$ 相似于 A^*. 但是可选取 α 使 $T+T^*$ 非奇异，这是因为，$T+T^*$ 非奇异，当且仅当 $T^{-1}(T+T^*)=I+T^{-1}T^*$ 非奇异，或当且仅当 $-1\notin\sigma(T^{-1}T^*)$. 但是，$T^{-1}T^*=e^{-2i\theta}S^{-1}S^*$，又因为可以选取 α 来得到任意 $\theta\in[0,2\pi]$，所以只需选取 θ 使 $-e^{2i\theta}\notin\sigma(S^{-1}S^*)$ 即可. 因而(b)蕴涵(c).

其次，假定(c)成立，并且写出 $R^{-1}AR=A^*$，其中 $R\in M_n$ 是非奇异 Hermite 矩阵. 于是 $R^{-1}A=A^*R^{-1}$ 且 $A=R(A^*R^{-1})$. 但 $(A^*R^{-1})^*=R^{-1}A=A^*R^{-1}$，因此 A 是两个 Hermite 矩阵 R 和 A^*R^{-1} 的乘积，且其中的 R 是非奇异的，因而(d)成立.

如果(d)成立，且 $A=HK$，其中 H 是非奇异矩阵，那么 $H^{-1}AH=KH=(HK)^*=A^*$，从而(b)成立. 如果 K 是非奇异矩阵，证明是类似的.

显然(d)蕴涵(e)；我们要证明(e)蕴涵(a). 如果 $A=HK$，其中 H 和 K 是 Hermite 矩阵，且都是奇异的，考虑 $U^*AU=(U^*HU)(U^*KU)$，其中，$U\in M_n$ 是酉矩阵，它把 H 对角化成形式

$$U^*HU=\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}=H',$$

使得 $D\in M_k$ 是非奇异对角矩阵，$k<n$. 把矩阵 U^*KU 划分成与 H' 同形的分块，使得

$$U^*AU=H'(U^*KU)=\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} K' & * \\ * & * \end{bmatrix}$$

$$=\begin{bmatrix} DK' & * \\ 0 & 0 \end{bmatrix}.$$

子块 $DK'\in M_k$ 是两个 Hermite 矩阵的乘积，其中有一个是非奇异的，于是根据(d)与(a)的等

价性，DK'相似于实矩阵 $B\in M_k$. 用 $J\in M_k$ 表示 B 的 Jordan 标准形，使得 A 相似于形如

$$C=\begin{bmatrix} J & * \\ 0 & 0 \end{bmatrix}$$

的矩阵 C. 矩阵 C 是上三角矩阵，且它的各特征值是 J 的各特征值再附加 $n-k$ 个零特征值，对于相应于任一非零特征值的诸子块，C 的 Jordan 标准形的 Jordan 块结构必须与 J 的相同，这是因为，如果 $\lambda\neq 0$，每个幂 $(C-\lambda I)^r$ 的(列)秩显然等于 $n-k+\operatorname{rank}(J-\lambda I)^r$，$r=1, 2, \cdots, n$. 特别是，$C$ 的相应于任一非特征值的 Jordan 块必须成共轭对出现，因而 C 的 Jordan 标准形相似于形如(3.4.6)的矩阵，它是一个实矩阵. □

173

习题

1. 证明 Hermite 矩阵的每个主子矩阵是 Hermite 矩阵. 这一性质对斜 Hermite 矩阵成立吗？对正规矩阵呢？

2. 如果 $A\in M_n$ 是 Hermite 矩阵，且 $S\in M_n$，证明 SAS^* 是 Hermite 矩阵. 如果 S 是非奇异的，SAS^{-1} 又怎样呢？

3. 设 $A, B\in M_n$ 是 Hermite 矩阵. 证明，A 与 B 相似，当且仅当它们酉相似. **提示**：如果 $A=SBS^{-1}$，证明 $A=U\Lambda U^*$ 和 $B=V\Lambda V^*$，其中 U 和 V 是酉矩阵，因而 $U^*AU=\Lambda=V^*BV$.

4. 验证(4.1.1)下面的性质 1～9.

5. 有时，通过证明一个矩阵与一个 Hermite 矩阵相似，可以证明它只有实特征值. 关于这方面的一个经典例子如下所述：设 $A=[a_{ij}]\in M_n(\mathbf{R})$ 是三对角矩阵；即 $a_{ij}=0$，若 $|i-j|>1$. 假定它的元素有很弱的对称性质，即对所有 $i=1, 2, \cdots, n-1$，有 $a_{i,i+1}a_{i+1,i}>0$. 证明，存在一个具有正对角元的实对角矩阵 D 使得 DAD^{-1} 是对称的. 因而得出 A 只有实特征值. 考察$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$，试说明为什么关于非对角元的符号假定是必不可少的. 试用极限理论证明，如果 $a_{i,i+1}a_{i+1,i}\geqslant 0$，那么特征值是实的结论仍然成立.

6. 证明在下述意义下每个矩阵 $A\in M_n$ 由它产生的 Hermite 型 x^*Ax 唯一确定：如果 $A=[a_{ij}]$，$B=[b_{ij}]\in M_n$ 是给定的，证明，对所有 $x\in\mathbf{C}^n$ 有 $x^*Ax=x^*Bx$，当且仅当 $A=B$. **提示**：如果对所有 $x\in\mathbf{C}^n$，$x^*Ax=0$，考虑 $(x+y)^*A(x+y)$，然后证明，对所有 $x, y\in\mathbf{C}^n$，有 $x^*Ay+y^*Ax=0$. 选取 $x=e_k$，$y=e^{i\theta}e_j$，$\theta\in\mathbf{R}$，证明，对所有 $\theta\in\mathbf{R}$ 和所有 $j, k=1, 2, \cdots, n$，有 $a_{kj}e^{2i\theta}=-a_{jk}$.

7. 证明，如果 $n\geqslant 2$，矩阵 $A\in M_n$ 不能由它产生的二次型 x^TAx 唯一确定；即：如果 $n\geqslant 2$，则存在 $A, B\in M_n$，且 $A\neq B$，使得对所有 $x\in\mathbf{C}^n$ 有 $x^TAx=x^TBx$. **提示**：如果 $C=-C^T$，x^TCx 是什么？

8. 证明，矩阵 $A\in M_n$ 不能由它产生的 Hermite 型的绝对值 $|x^*Ax|$ 所确定. **提示**：设 $A=\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$，然后证明，对所有 $x\in\mathbf{C}^2$，$|x^*Ax|=|x^*A^Tx|$.

174

9. 证明，在下述意义下，矩阵 A 几乎由它生成的 Hermite 双线性型的绝对值所确定：若

A, $B\in M_n$ 是给定的，证明，对所有 x, $y\in \mathbf{C}^n$ 有 $|x^*Ay|=|x^*By|$，当且仅当对某个 $\theta\in\mathbf{R}$ 有 $A=e^{i\theta}B$. **提示**：设 $A=[a_{ij}]$ 且 $B=[b_{ij}]$. 利用 $x=e_i$ 和 $y=e_j$ 证明，对所有 i, $j=1$, $\cdots$, n 有 $|a_{ij}|=|b_{ij}|$. 设 $x=e_i$, $y=se_j+te_k$，证明 $|sa_{ij}+ta_{ik}|^2=|sb_{ij}+tb_{ik}|^2$，因而对所有 s, $t\in\mathbf{C}$ 有 $\mathrm{Re}(s\bar{t}[a_{ij}\bar{a}_{ik}-b_{ij}\bar{b}_{ik}])=0$. 若 $b_{ij}b_{ik}\neq 0$，推出 $a_{ij}/b_{ij}=a_{ik}/b_{ik}$.

10. 证明，$A\in M_n$ 是 Hermite 矩阵，当且仅当 iA 是斜 Hermite 矩阵. 证明，一个斜 Hermite 矩阵的各特征值都是纯虚的，而一个斜 Hermite 矩阵的平方的各特征值是非正实数.

11. 如果 A, $B\in M_n$ 是 Hermite 矩阵，证明 $\mathrm{tr}(AB)^2\leqslant \mathrm{tr}\, A^2B^2$. **提示**：证明 $AB-BA$ 是斜 Hermite 矩阵，然后考虑 $\mathrm{tr}(AB-BA)^2$.

12. 如果 $A\in M_n$ 是 Hermite 矩阵，证明，A 的秩等于 A 的非零特征值的个数，但是对于非 Hermite 矩阵一般不成立. **提示**：$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$.

13. 若 $A\in M_n$ 是 Hermite 矩阵且 $A\neq 0$，证明

$$\mathrm{rank}(A)\geqslant\frac{[\mathrm{tr}\,A]^2}{\mathrm{tr}\,A^2}$$

上述等式成立当且仅当存在一个具有标准正交列的矩阵 $U=[u_1\cdots u_r]\in M_{n,r}$ 和某个 $a\in\mathbf{R}$ 使得 $A=aUU^*$；即 A 是一个酉射影的实纯量倍数. **提示**：若 λ_1, $\cdots$, λ_r 是 A 的非零特征值，Cauchy-Schwarz 不等式是说：

$$[\mathrm{tr}\,A]^2=\Big(\sum_{i=1}^{r}\lambda_i\Big)^2\leqslant r\sum_{i=1}^{r}\lambda_i^2=r\,\mathrm{tr}\,A^2,$$

其中等式成立当且仅当所有 λ_i 相等.

14. 斜 Hermite 矩阵 $A\in M_n$ 满足恒等式 $A=-A^*$. 如果 $\theta\in\mathbf{R}$，证明，$A=e^{i\theta}A^*$，当且仅当 $e^{-i\theta/2}A$ 是 Hermite 矩阵. 对于 $\theta=\pi$，这是什么？对 $\theta=0$ 呢？说明为什么斜 Hermite 矩阵类可以看作"广义 Hermite"矩阵的无限多个类中的一个，并且描述每一个这样的矩阵类的结构.

15. 设 Hermite 矩阵 $A=[a_{ij}]\in M_n$ 已写成如下的分块形式：

$$A=\begin{bmatrix}a_{11} & x^*\\ x & \tilde{A}\end{bmatrix},$$

其中，$x\in\mathbf{C}^{n-1}$，$\tilde{A}\in M_{n-1}$. 证明

$$\det A=a_{11}\det\tilde{A}-x^*(\mathrm{adj}\,\tilde{A})x$$

175

其中 $\mathrm{adj}\,\tilde{A}$ 是 $\tilde{A}$ 的经典伴随[见(0.8.2)]. 使 A 满足更弱的什么条件就能保证这个公式成立？**提示**：按 A 的第 1 列进行行列式的 Lanlace 余子式展开(0.3.1)，然后对所得到诸余子式的第一行进行展开.

4.2 Hermite 矩阵的特征值的变分特征

对于一般的矩阵 $A\in M_n$，关于特征值的唯一特征是，它们是特征方程 $p_A(t)=0$ 的根. 但是，对于 Hermite 矩阵，特征值可以表征为一系列最优化问题的解.

因为 Hermite 矩阵 $A\in M_n$ 的特征值都是实数，我们将约定，它们将按其大小递增(不减)顺序排列：

$$\lambda_{\min}=\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_{n-1}\leqslant\lambda_n=\lambda_{\max}. \tag{4.2.1}$$

很容易把最小的和最大的特征值描述成约束最小值和约束最大值问题. 下述变分特征定理冠上了两个英国物理学家的名字，并且称 x^*Ax/x^*x 为 Rayleigh-Ritz 比.

4.2.2 定理(Rayleigh-Ritz) 设 $A\in M_n$ 是 Hermite 矩阵，且设 A 的诸特征值取(4.2.1)中的顺序，那么

$$\lambda_1x^*x\leqslant x^*Ax\leqslant\lambda_nx^*x \text{ 对所有 } x\in\mathbf{C}^n \text{ 成立；}$$

$$\lambda_{\max}=\lambda_n=\max_{x\neq0}\frac{x^*Ax}{x^*x}=\max_{x^*x=1}x^*Ax,$$

$$\lambda_{\min}=\lambda_1=\min_{x\neq0}\frac{x^*Ax}{x^*x}=\min_{x^*x=1}x^*Ax.$$

证明：因为 A 是 Hermite 矩阵，所以存在酉矩阵 $U\in M_n$ 使得 $A=U\Lambda U^*$，其中 $\Lambda=\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$. 对任一 $x\in\mathbf{C}^n$，有

$$x^*Ax=x^*U\Lambda U^*x=(U^*x)^*\Lambda(U^*x)$$

$$=\sum_{i=1}^n\lambda_i\,|(U^*x)_i|^2$$

由于每项 $|(U^*x)_i|^2$ 是非负的，于是有

176

$$\lambda_{\min}\sum_{i=1}^n|(U^*x)_i|^2\leqslant x^*Ax=\sum_{i=1}^n\lambda_i\,|(U^*x)_i|^2\leqslant\lambda_{\max}\sum_{i=1}^n|(U^*x)_i|^2.$$

因为 U 是酉矩阵，所以

$$\sum_{i=1}^n|(U^*x)_i|^2=\sum_{i=1}^n|x_i|^2=x^*x,$$

因而我们已经证明，

$$\lambda_1x^*x=\lambda_{\min}x^*x\leqslant x^*Ax\leqslant\lambda_{\max}x^*x=\lambda_nx^*x. \tag{4.2.3}$$

因为，如果 x 是 A 的相应于特征值 λ_1 的特征向量，那么 $x^*Ax=x^*\lambda_1x=\lambda_1x^*x$，且对于 λ_n 也有类似的结果，因此，这些不等式可取等式. 其余的论断容易从(4.2.3)推出. 如果 $x\neq0$，则有

$$\frac{x^*Ax}{x^*x}\leqslant\lambda_n,$$

当 x 是 A 的相应于 λ_n 的特征向量时，等式成立，因而

$$\max_{x\neq0}\frac{x^*Ax}{x^*x}=\lambda_n. \tag{4.2.4}$$

最后，如果 $x\neq0$，则有

$$\frac{x^*Ax}{x^*x}=\left(\frac{x}{\sqrt{x^*x}}\right)^*A\left(\frac{x}{\sqrt{x^*x}}\right)\text{和}\left(\frac{x}{\sqrt{x^*x}}\right)^*\left(\frac{x}{\sqrt{x^*x}}\right)=1.$$

因此，条件(4.2.4)等价于条件

$$\max_{x^*x=1}x^*Ax=\lambda_n. \tag{4.2.5}$$

关于 λ_1 的论证是类似的. □

(4.2.5)的几何解释是，当 x 取遍 $\mathbf{C}^n$ 中的单位球(它是紧集)时，函数 x^*Ax 的最大值是 λ_n. x^*Ax 的取值范围(4.2.3)则给出了下述特征值包含结论.

4.2.6 推论 设 $A\in M_n$ 是一个给定的 Hermite 矩阵，$x\in\mathbf{C}^n$ 是一个给定的非零向量，又设 $\alpha\equiv x^*Ax/x^*x$. 那么，在区间$(-\infty, \alpha]$中至少有 A 的一个特征值，同样在区间$[\alpha, \infty)$中也至少有 A 的一个特征值.

练习 证明推论(4.2.6).

练习 关于 λ_1 的类似于(4.2.5)的结论及其几何解释是什么?

Rayleigh-Ritz 定理给出了 Hermite 矩阵 A 的最大特征值和最小特征值的一个变分特征，然而，其余的特征值又怎样呢? 假定 $A=U\Lambda U^*$，且 $U=[u_1, u_2, \cdots, u_n]$；$U$ 的诸列是 A 的标准正交向量. 如果只考虑与 u_1 正交的那些向量 $x\in\mathbf{C}^n$，那么，对在(4.2.2)证明中的主要恒等式，应做如下修正：

177

$$x^*Ax=\sum_{i=1}^n\lambda_i|(U^*x)_i|^2=\sum_{i=1}^n\lambda_i|u_i^*x|^2=\sum_{i=2}^n\lambda_i|u_i^*x|^2.$$

这是 $\lambda_2, \lambda_3, \cdots, \lambda_n$ 的一个凸组合，因而，只要 x 与 U 的第 1 个列向量正交，就有

$$x^*Ax=\sum_{i=2}^n\lambda_i|u_i^*x|^2\geqslant\lambda_2\sum_{i=2}^n|u_i^*x|^2=\lambda_2\sum_{i=1}^n|(U^*x)_i|^2=\lambda_2x^*x.$$

如果取 $x=u_2$，这个不等式就变成等式，于是，有第二个最小特征值的一个特征

$$\min_{\substack{x\neq 0\\ x\perp u_1}}\frac{x^*Ax}{x^*x}=\min_{\substack{x^*x=1\\ x\perp u_1}}x^*Ax=\lambda_2. \tag{4.2.7}$$

练习 试推广上述论证，进而证明

$$\min_{\substack{x\neq 0\\ x\perp u_1,u_2,\cdots,u_{k-1}}}\frac{x^*Ax}{x^*x}=\min_{\substack{x^*x=1\\ x\perp u_1,u_2,\cdots,u_{k-1}}}x^*Ax=\lambda_k,\quad k=2,3,\cdots,n. \tag{4.2.8}$$

练习 证明

$$\max_{\substack{x\neq 0\\ x\perp u_n,u_{n-1},\cdots,u_{n-k+1}}}\frac{x^*Ax}{x^*x}=\max_{\substack{x^*x=1\\ x\perp u_n,u_{n-1},\cdots,u_{n-k+1}}}x^*Ax=\lambda_{n-k},\quad k=1,2,\cdots,n-1. \tag{4.2.9}$$

遗憾的是，这些公式的实用价值不大，这是因为它们要求明确知道某些特征向量，而通常又没有这方面的信息. 但是，(4.2.7)以及相关的公式(4.2.8)和(4.2.9)可以作为推导出一个有用的特征的出发点. 设 $w\in\mathbf{C}^n$ 是给定的向量. 那么

$$\sup_{\substack{x^*x=1\\ x\perp w}}x^*Ax=\sup_{\substack{x^*x=1\\ x\perp w}}x^*U\Lambda U^*x=\sup_{\substack{x^*x=1\\ x\perp w}}\sum_{i=1}^n\lambda_i|(U^*x)_i|^2$$

$$=\sup_{\substack{z^*z=1\\ x=Uz\perp w}}\sum_{i=1}^n\lambda_i|z_i|^2=\sup_{\substack{z^*z=1\\ z\perp U^*w}}\sum_{i=1}^n\lambda_i|z_i|^2$$

$$\geqslant\sup_{\substack{z^*z=1\\ z\perp U^*w\\ z_1=z_2=\cdots=z_{n-2}=0}}\sum_{i=1}^n\lambda_i|z_i|^2$$

$$= \sup_{\substack{|z_{n-1}|^2+|z_n|^2=1 \\ z \perp U^* w}} \lambda_{n-1}|z_{n-1}|^2 + \lambda_n |z_n|^2 \geqslant \lambda_{n-1}. \tag{4.2.10}$$

178

在上述第二行论证中，令 $z=U^*x$，且用到了以下事实：如果 $x^*x=1$，那么 U 是酉矩阵可推出 $z^*z=1$. 上述第一个不等式从以下事实得来的：如果缩小某集合，然后在其上取上确界，则上确界的值不会增加. 最后一个不等式是从次序关系 $\lambda_n \geqslant \lambda_{n-1}$ 以及常用的凸组性质得来的.

在上述论证中，向量 w 是固定的，但又是任意的，因而可以在所有 w 上取(4.2.10)的下确界，这便得到

$$\inf_{w\in \mathbf{C}^n} \sup_{\substack{x^*x=1 \\ x \perp w}} x^*Ax \geqslant \lambda_{n-1}.$$

但是，如果 $w=u_n$，则(4.2.10)中等式成立，因此有

$$\inf_{w\in \mathbf{C}^n} \sup_{\substack{x^*x=1 \\ x \perp w}} x^*Ax = \lambda_{n-1},$$

这个在形式上比(4.2.7)稍微复杂一些的特征并不需要知道 A 的任何特征向量. 因为 $x=u_{n-1}$ 时有 $x^*Ax=\lambda_{n-1}$，所以常常看到这个公式用“max”代替“sup”，用“min”代替“inf”. 这是下述 Courant-Fischer“极小一极大定理”的基本思想.

4.2.11 定理(Courant-Fischer) 设 $A\in M_n$ 是具有特征值 $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ 的 Hermite 矩阵，k 是给定的整数，$1\leqslant k\leqslant n$，那么

$$\min_{w_1,w_2,\cdots,w_{n-k}\in \mathbf{C}^n} \max_{\substack{x\neq 0, x\in \mathbf{C}^n \\ x\perp w_1,w_2,\cdots,w_{n-k}}} \frac{x^*Ax}{x^*x} = \lambda_k, \tag{4.2.12}$$

$$\max_{w_1,w_2,\cdots,w_{k-1}\in \mathbf{C}^n} \min_{\substack{x\neq 0, x\in \mathbf{C}^n \\ x\perp w_1,w_2,\cdots,w_{k-1}}} \frac{x^*Ax}{x^*x} = \lambda_k. \tag{4.2.13}$$

附注 如果在(4.2.12)中 $k=n$，或在(4.2.13)中 $k=1$，约定去掉在外侧取极小或极大过程，因为这时取极小或极大的集合是空集. 在这两种情形，上述结论简化为 Rayleigh-Ritz 定理(4.2.2).

证明： 我们只考虑(4.2.12)，而(4.2.13)的证法是类似的. 把 A 写成 $A=U\Lambda U^*$，其中 U 是酉矩阵，而 $\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$，且设 $1<k\leqslant n$. 如果 $x\neq 0$，则

179

$$\frac{x^*Ax}{x^*x} = \frac{(U^*x)^*\Lambda(U^*x)}{x^*x} = \frac{(U^*x)^*\Lambda(U^*x)}{(U^*x)^*(U^*x)},$$

且 $\{U^*x: x\in \mathbf{C}^n \text{ 且 } x\neq 0\}=\{y\in \mathbf{C}^n: y\neq 0\}$. 因此，如果给定 $w_1, w_2, \cdots, w_{n-k}\in \mathbf{C}^n$，便有

$$\sup_{\substack{x\neq 0 \\ x\perp w_1,\cdots,w_{n-k}}} \frac{x^*Ax}{x^*x} = \sup_{\substack{y\neq 0 \\ y\perp U^*w_1,\cdots,U^*w_{n-k}}} \frac{y^*\Lambda y}{y^*y}$$

$$= \sup_{\substack{y^*y=1 \\ y\perp U^*w_1,\cdots,U^*w_{n-k}}} \sum_{i=1}^{n} \lambda_i |y_i|^2$$

$$\geqslant \sup_{\substack{y^*y=1 \\ y\perp U^*w_1,\cdots,U^*w_{n-k} \\ y_1=y_2=\cdots=y_{k-1}=0}} \sum_{i=k}^{n} \lambda_i |y_i|^2$$

$$= \sup_{\substack{|y_k|^2+|y_{k+1}|^2+\cdots+|y_n|^2=1 \\ y\perp U^*w_1,\cdots,U^*w_{n-k}}} \sum_{i=k}^{n} \lambda_i |y_i|^2 \geqslant \lambda_k.$$

这说明，对任意 $n-k$ 个向量 $w_1, w_2, \cdots, w_{n-k}$，

$$\sup_{\substack{x\neq 0 \\ x\perp w_1,\cdots,w_{n-k}}} \frac{x^*Ax}{x^*x} \geqslant \lambda_k.$$

而(4.2.9)说明，为使等式成立，诸向量 w_i 总有一种选择，即 $w_i=u_{n-i+1}$，其中 $U=[u_1\cdots u_n]$. 因此，

$$\inf_{w_1,\cdots,w_{n-k}} \sup_{\substack{x\neq 0 \\ x\perp w_1,\cdots,w_{n-k}}} \frac{x^*Ax}{x^*x} = \lambda_k,$$

又因为极值是可以到达的，故可以用“min”和“max”代替“inf”和“sup”. 关于(4.2.13)的证明是类似的. □

练习 给出(4.2.13)的详细证明.

习题

1. 设 $A\in M_n$ 是具有特征值 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$ 的 Hermite 矩阵，试利用(4.2.11)证明

$$\lambda_k = \min_{S_k\subset \mathbf{C}^n} \max_{\substack{x\neq 0 \\ x\in S_k}} \frac{x^*Ax}{x^*x}, \quad k=1,2,\cdots,n,$$

$$\lambda_k = \max_{S_{n-k+1}\subset \mathbf{C}^n} \min_{\substack{x\neq 0 \\ x\in S_{n-k+1}}} \frac{x^*Ax}{x^*x}, \quad k=1,2,\cdots,n,$$

180

在这两个式子中，S_j 表示 j 维子空间，其中在外侧取极小和取极大是对所有指定维数的子空间而言的.

2. 如果 $A\in M_n$ 是 Hermite 矩阵，证明下述三个求极大值的问题有相同的解：

(a) $\max\limits_{x^*x=1} x^*Ax$,

(b) $\max\limits_{x\neq 0} \dfrac{x^*Ax}{x^*x}$,

(c) $\max\limits_{x^*Ax=1} \dfrac{1}{x^*x}$，如果 A 至少有一个特征值是正数.

3. 如果 $A\in M_n$ 是 Hermite 矩阵，且 $x^*x=1$，证明

$$\lambda_{\max} \geqslant x^*Ax \geqslant \lambda_{\min}.$$

4. 通过考察 $A=\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$来说明，在(4.2.2)中，关于 A 是 Hermite 矩阵的假定必不可少. $\max\{x^TAx/x^Tx:\ 0\neq x\in \mathbf{R}^2\}$等于什么？$\max \mathrm{Re}\{x^*Ax/x^*x:\ 0\neq x\in\mathbf{C}^2\}$等于什么？

5. 设 $A\in M_n$ 有特征值$\{\lambda_i\}$. 证明，即使 A 是 Hermite 矩阵，其特征值也有如下界限：

$$\min_{x\neq 0}\left|\frac{x^*Ax}{x^*x}\right| \leqslant |\lambda_i| \leqslant \max_{x\neq 0}\left|\frac{x^*Ax}{x^*x}\right|, \quad i=1,2,\cdots,n.$$

提示：取 x 为 A 的一个特征向量，并且用 $A=\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$说明上、下两个界都是可以达到的.

4.3　变分特征的某些应用

在 Courant-Fischer 定理的许多重要应用中，最简单的应用是关于 $A+B$ 的诸特征值与 A 的特征值的比较问题. 我们用 $\lambda_i(A)$ 表示矩阵 A 的诸特征值.

4.3.1 定理(Weyl)　设 A、$B\in M_n$ 是 Hermite 矩阵. 又设诸特征值 $\lambda_i(A)$、$\lambda_i(B)$ 以及 $\lambda_i(A+B)$ 均按递增顺序(4.2.1)排列. 则对每个 $k=1,2,\cdots,n$，有

$$\lambda_k(A)+\lambda_1(B)\leqslant\lambda_k(A+B)\leqslant\lambda_k(A)+\lambda_n(B).\tag{4.3.2}$$

证明：对于任意非 $x\in\mathbf{C}^n$，有不等式

$$\lambda_1(B)\leqslant\frac{x^*Bx}{x^*x}\leqslant\lambda_n(B),$$

[181] 因而对任意 $k=1,2,\cdots,n$，有

$$\begin{aligned}\lambda_k(A+B)&=\min_{w_1,\cdots,w_{n-k}\in\mathbf{C}^n}\ \max_{\substack{x\neq0\\x\perp w_1,\cdots,w_{n-k}}}\frac{x^*(A+B)x}{x^*x}\\&=\min_{w_1,\cdots,w_{n-k}\in\mathbf{C}^n}\ \max_{\substack{x\neq0\\x\perp w_1,\cdots,w_{n-k}}}\left[\frac{x^*Ax}{x^*x}+\frac{x^*Bx}{x^*x}\right]\\&\geqslant\min_{w_1,\cdots,w_{n-k}\in\mathbf{C}^n}\ \max_{\substack{x\neq0\\x\perp w_1,\cdots,w_{n-k}}}\left[\frac{x^*Ax}{x^*x}+\lambda_1(B)\right]=\lambda_k(A)+\lambda_1(B).\end{aligned}$$

类似的论证同样可以确定它的上界. □

练习　说明在(4.3.1)中给出的界中可以取等号. **提示**：设 $\{u_1,u_2,\cdots,u_n\}$ 是由 A 的特征向量组成的标准正交组，且 $Au_i=\lambda_i(A)u_i$. 分别对 $\alpha>0$ 和 $\alpha<0$ 考察 $B=\alpha u_iu_i^*$.

Weyl 定理对任意 Hermite 矩阵 A 和 B 给出了 $A+B$ 的特征值的上界和下界. 如果限制 B 具有一种特殊的形式，例如 B 为正定矩阵、秩 1 矩阵、秩 k 矩阵或加边矩阵，还可以得到更为精细的表述.

矩阵 $B\in M_n$ 称为半正定的，是指它对所有 $x\in\mathbf{C}^n$ 有 $x^*Bx\geqslant0$. 一个等价的条件是，B 是 Hermite 矩阵，且所有特征值都是非负的(见第 7 章). 下述结果是 Weyl 定理的直接推论，称之为单调性定理，它说明，如果一个 Hermite 矩阵加上一个半正定矩阵，则它的所有特征值都会增加.

4.3.3 推论　设 A，$B\in M_n$ 是 Hermite 矩阵. 假定 B 是半正定矩阵，且 A 和 $A+B$ 的诸特征值均按递增顺序(4.2.1)排列. 则对所有 $k=1,2,\cdots,n$，

$$\lambda_k(A)\leqslant\lambda_k(A+B).$$

证明：利用(4.3.2)中的下界及 $\lambda_1(B)\geqslant0$ 的事实. □

如果矩阵 B 有秩 1，那么用 A 的特征值作为 $A+B$ 的特征值的界具有交错定理的形式：在 $A+B$ 的每对相邻的单号数(或双号数)特征值间至少存在 A 的一个特征值.

4.3.4 定理　设 $A\in M_n$ 是 Hermite 矩阵，$z\in\mathbf{C}^n$ 是给定的向量. 如果 A 和 $A\pm zz^*$ 的诸特征值
[182] 均按递增顺序(4.2.1)排列，则有

(a) $\lambda_k(A\pm zz^*)\leqslant\lambda_{k+1}(A)\leqslant\lambda_{k+2}(A\pm zz^*)$,　$k=1,2,\cdots,n-2$,

(b) $\lambda_k(A)\leqslant\lambda_{k+1}(A+zz^*)\leqslant\lambda_{k+2}(A)$,　$k=1,2,\cdots,n-2$.

证明：设 $1\leqslant k\leqslant n-2$，利用(4.2.12)可导出下述关系：

$$\begin{aligned}
\lambda_{k+2}(A\pm zz^*) &= \min_{w_1,\cdots,w_{n-k-2}\in\mathbf{C}^n}\ \max_{\substack{x\neq 0\\ x\perp w_1,\cdots,w_{n-k-2}}} \frac{x^*(A\pm zz^*)x}{x^*x}\\
&\geqslant \min_{w_1,\cdots,w_{n-k-2}\in\mathbf{C}^n}\ \max_{\substack{x\neq 0\\ x\perp w_1,\cdots,w_{n-k-2}\\ x\perp z}} \frac{x^*(A\pm zz^*)x}{x^*x}\\
&= \min_{\substack{w_1,\cdots,w_{n-k-2}\in\mathbf{C}^n\\ w_{n-k-1}=z}}\ \max_{\substack{x\neq 0\\ x\perp w_1,\cdots,w_{n-k-2},w_{n-k-1}}} \frac{x^*Ax}{x^*x}\\
&\geqslant \min_{w_1,\cdots,w_{n-k-2},w_{n-k-1}\in\mathbf{C}^n}\ \max_{\substack{x\neq 0\\ x\perp w_1,\cdots,w_{n-k-2},w_{n-k-1}}} \frac{x^*Ax}{x^*x}\\
&= \lambda_{k+1}(A).
\end{aligned}$$

现在设 $2\leqslant k\leqslant n-1$，利用(4.2.13)可以导出下述关系：

$$\begin{aligned}
\lambda_k(A\pm zz^*) &= \max_{w_1,\cdots,w_{k-1}\in\mathbf{C}^n}\ \min_{\substack{x\neq 0\\ x\perp w_1,\cdots,w_{k-1}}} \frac{x^*(A\pm zz^*)x}{x^*x}\\
&\leqslant \max_{w_1,\cdots,w_{k-1}\in\mathbf{C}^n}\ \min_{\substack{x\neq 0\\ x\perp w_1,\cdots,w_{k-1}\\ x\perp z}} \frac{x^*(A\pm zz^*)x}{x^*x}\\
&= \max_{\substack{w_1,\cdots,w_{k-1}\in\mathbf{C}^n\\ w_k=z}}\ \min_{\substack{x\neq 0\\ x\perp w_1,\cdots,w_{k-1},w_k}} \frac{x^*Ax}{x^*x}\\
&\leqslant \max_{w_1,\cdots,w_{k-1},w_k\in\mathbf{C}^n}\ \min_{\substack{x\neq 0\\ x\perp w_1,\cdots,w_{k-1},w_k}} \frac{x^*Ax}{x^*x} = \lambda_{k+1}(A).
\end{aligned}$$

合并这两组不等式，由此得到所要求的不等式组. □

如果 $B\in M_n$ 是一个 Hermite 矩阵，且 $B=U\Lambda U^*$，其中，$U=[u_1u_2\cdots u_n]$是酉矩阵，$\Lambda=\mathrm{diag}(\beta_1,\beta_2,\cdots,\beta_n)$，则 B 的秩等于非零特征值的个数. 如果 B 的秩小于或等于 r，则可假定 $\beta_{r+1}=\cdots=\beta_n=0$. 如果秩小于 r，则 $\beta_1,\beta_2,\cdots,\beta_r$ 含有某些为零的特征值，表示式 [183]

$$B=\sum_{i=1}^{r}\beta_i u_i u_i^* \tag{4.3.5}$$

是表示 $B=U\Lambda U^*$ 的另一种形式. 反过来，形如(4.3.5)的任何矩阵，只要所有 $\beta_i\neq 0$ 且$\{u_i\}$是无关组，那么它就有秩 r；如果不知道$\{u_i\}$是否无关，那么 B 的秩至多是 r. Weyl 的下一个定理给出了当 B 有秩 r 时 $A+B$ 的诸特征值的界，它来源于积分方程理论. 本定理是(4.3.4)的秩 1 情形的简单推广.

4.3.6 定理　设 A，$B\in M_n$ 是 Hermite 矩阵，并且假定 B 至多有秩 r. 则

(a) $\lambda_k(A+B)\leqslant\lambda_{k+r}(A)\leqslant\lambda_{k+2r}(A+B)$,　$k=1,2,\cdots,n-2r$;

(b) $\lambda_k(A)\leqslant\lambda_{k+r}(A+B)\leqslant\lambda_{k+2r}(A)$,　$k=1,2,\cdots,n-2r$;

(c) 如果 $A=U\Lambda U^*$，其中，$U=[u_1 u_2 \cdots u_n]\in M_n$ 是酉矩阵，$\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，且 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$，又如果

$$B = \lambda_n u_n u_n^* + \lambda_{n-1} u_{n-1} u_{n-1}^* + \cdots + \lambda_{n-r+1} u_{n-r+1} u_{n-r+1}^*,$$

则 $\lambda_{\max}(A-B)=\lambda_{n-r}(A)$.

证明：设 $B=\alpha_1 u_1 u_1^* +\cdots+\alpha_r u_r u_r^*$，其中集合 $\{u_1, \cdots, u_r\}\subset\mathbf{C}^n$ 不一定无关. (a)和(b)的证明恰好与(4.3.4)中(a)和(b)的证明相同. 只是在前面放上一个条件"$x\perp z$"的地方，现在加上 r 个条件"$x\perp u_1, \cdots, u_r$"，并完成相应的论证. 关于(c)，注意到 $u_1, \cdots, u_n$ 都是 $A-B$ 的特征向量，但是对于 $k=n-r+1, n-r+2, \cdots, n$，$(A-B)u_k=0$，而对于 $k=1, 2, \cdots, n-r$，$(A-B)u_k=\lambda_k u_k$. 因为 $\lambda_{n-r}\geqslant\lambda_{n-r-1}\geqslant\cdots\geqslant\lambda_1$，所以 $A-B$ 的最大特征值是 λ_{n-r}. □

练习 给出(4.3.6)中的(a)和(b)的详细证明.

上述结果所提供的足够信息，使我们能够推导出 Weyl 关于两个 Hermite 矩阵之和的特征值的下述一般结果.

4.3.7 定理(Weyl) 设 $A, B\in M_n$ 是 Hermite 矩阵，且 A，B 和 $A+B$ 的诸特征值按(4.2.1)的递增顺序排列，那么，对于使得 $1\leqslant j, k\leqslant n$ 且 $j+k\geqslant n+1$ 的每对整数 j, k，有

184

$$\lambda_{j+k-n}(A+B) \leqslant \lambda_j(A)+\lambda_k(B),$$

而对于使得 $1\leqslant j, k\leqslant n$ 和 $j+k\leqslant n+1$ 的每对整数 j, k，有

$$\lambda_j(A)+\lambda_k(B) \leqslant \lambda_{j+k-1}(A+B).$$

证明：设 j, k 是满足第一组条件的已知整数. 设 $A=U\Lambda(A)U^*$，$B=V\Lambda(B)V^*$，其中，$U=[u_1 u_2 \cdots u_n]\in M_n$ 和 $V=[v_1 v_2 \cdots v_n]\in M_n$ 是酉矩阵，$\Lambda(A)=\mathrm{diag}(\lambda_1(A), \cdots, \lambda_n(A))\in M_n$ 且 $\Lambda(B)=\mathrm{diag}(\lambda_1(B), \cdots, \lambda_n(B))\in M_n$，因而

$$A_j \equiv \lambda_n(A)u_n u_n^* + \lambda_{n-1}(A)u_{n-1}u_{n-1}^* + \cdots + \lambda_{j+1}(A)u_{j+1}u_{j+1}^*$$

的秩至多为 $n-j$，$B_k\equiv\lambda_n(B)v_n v_n^* +\cdots+\lambda_{k+1}(B)v_{k+1}v_{k+1}^*$ 的秩至多为 $n-k$，而 A_j+B_k 的秩至多为 $2n-j-k$. 于是，根据(4.3.6c)，$\lambda_n(A-A_j)=\lambda_j(A)$，$\lambda_n(B-B_k)=\lambda_k(B)$，再根据(4.3.6b)，$\lambda_n(A-A_j+B-B_k)=\lambda_n(A+B-(A_j+B_k))\geqslant\lambda_{n-(2n-j-k)}(A+B)=\lambda_{j+k-n}(A+B)$(其中 $k+r=n$ 且 $r=2n-j-k$). 另外根据(4.3.2)有

$$\lambda_n(A-A_j+B-B_k)\leqslant\lambda_n(A-A_j)+\lambda_n(B-B_k),$$

(其中 $k=n$). 因此，有

$$\begin{aligned}\lambda_j(A)+\lambda_k(B) &= \lambda_n(A-A_j)+\lambda_n(B-B_k)\geqslant\lambda_n(A-A_j+B-B_k)\\ &=\lambda_n((A+B)-(A_j+B_k))\geqslant\lambda_{j+k-n}(A+B),\end{aligned}$$

这就是要证的第一组不等式. 第二组不等式可直接由第一组不等式推出，只需把它应用于 $-A$ 和 $-B$ 即可. □

练习 试给出从(4.3.7)中的第一组不等式推导出第二组不等式的详细证明. **提示**：利用(4.3.7)得到 $\lambda_{j+k-n}(-A-B)$ 的一个上界，然后利用下面的事实：如果 $A\in M_n$ 是 Hermite 矩阵，则 $\lambda_i(-A)=-\lambda_{n-i+1}(A)$.

作为这种类型的最后一个结果，考虑关于 $A+B$ 的特征值的交错定理，其中，假定 A 和 B 中的每一个都具有特殊的形式. 这个结果类似于(4.3.4)中假定 B 有秩 1 的情形，称之为加边

矩阵的交错特征值定理.

4.3.8 定理 设 $A\in M_n$ 是给定的 Hermite 矩阵，$y\in\mathbf{C}^n$ 是给定的向量，$a\in\mathbf{R}$ 是给定的实数. 设 $\hat{A}\in M_{n+1}$ 是在 A 旁镶上 y 和 a 后得到的 Hermite 矩阵，如下所述：

$$\hat{A}=\begin{bmatrix}A & y\\ y^* & a\end{bmatrix}.$$

[185]

设 A 和 $\hat{A}$ 的特征值分别用 $\{\lambda_i\}$ 和 $\{\hat{\lambda}_i\}$ 表示，且假定它们按递增顺序 $\lambda_1\leqslant\cdots\leqslant\lambda_n$ 和 $\hat{\lambda}_1\leqslant\hat{\lambda}_2\leqslant\cdots\leqslant\hat{\lambda}_n\leqslant\hat{\lambda}_{n+1}$ 排列，那么

$$\hat{\lambda}_1\leqslant\lambda_1\leqslant\hat{\lambda}_2\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_{n-1}\leqslant\hat{\lambda}_n\leqslant\lambda_n\leqslant\hat{\lambda}_{n+1}. \tag{4.3.9}$$

证明：设 k 是给定的整数，且 $1\leqslant k\leqslant n$. 我们要证明 $\hat{\lambda}_k\leqslant\lambda_k\leqslant\hat{\lambda}_{k+1}$. 设 $\hat{x}=[x^T\zeta]^T\in\mathbf{C}^{n+1}$，$x\in\mathbf{C}^n$，$\zeta\in\mathbf{C}$，且 $\hat{w}_i=[w_i^T w]^T\in\mathbf{C}^{n+1}$，$w_i\in\mathbf{C}^n$，$w\in\mathbf{C}$. 利用 Courant-Fischer 定理的 (4.2.12) 可导出

$$\begin{aligned}
\hat{\lambda}_{k+1}&=\min_{w_1,\cdots,w_{(n+1)-(k+1)}\in\mathbf{C}^{n+1}}\ \max_{\substack{\hat{x}\neq 0,\ \hat{x}\in\mathbf{C}^{n+1}\\ \hat{x}\perp w_1,\cdots,w_{(n+1)-(k+1)}}}\frac{\hat{x}^*\hat{A}\hat{x}}{\hat{x}^*\hat{x}}\\
&\geqslant\min_{\hat{w}_1,\cdots,\hat{w}_{n-k}\in\mathbf{C}^{n+1}}\ \max_{\substack{\hat{x}\neq 0\\ \hat{x}\perp\hat{w}_1,\cdots,\hat{w}_{n-k}\\ \hat{x}\perp e_{n+1}}}\frac{\hat{x}^*\hat{A}\hat{x}}{\hat{x}^*\hat{x}}\\
&=\min_{w_1,\cdots,w_{n-k}\in\mathbf{C}^n}\ \max_{\substack{x\neq 0,\ x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{n-k}}}\frac{x^*Ax}{x^*x}=\lambda_k.
\end{aligned}$$

关于 λ_k 的下界，利用(4.2.13).

$$\begin{aligned}
\hat{\lambda}_k&=\max_{w_1,\cdots,w_{k-1}\in\mathbf{C}^{n+1}}\ \min_{\substack{\hat{x}\neq 0,\ \hat{x}\in\mathbf{C}^{n+1}\\ \hat{x}\perp w_1,\cdots,w_{k-1}}}\frac{\hat{x}^*\hat{A}\hat{x}}{\hat{x}^*\hat{x}}\\
&\leqslant\max_{\hat{w}_1,\cdots,\hat{w}_{k-1}\in\mathbf{C}^{n+1}}\ \min_{\substack{\hat{x}\neq 0\\ \hat{x}\perp\hat{w}_1,\cdots,\hat{w}_{k-1}\\ \hat{x}\perp e_{n+1}}}\frac{\hat{x}^*\hat{A}\hat{x}}{\hat{x}^*\hat{x}}\\
&=\max_{w_1,\cdots,w_{k-1}\in\mathbf{C}^n}\ \min_{\substack{x\neq 0,\ x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{k-1}}}\frac{x^*Ax}{x^*x}=\lambda_k.
\end{aligned}$$

□

我们已经看到关于特征值的交错定理的两个例子：如果给某个 Hermite 矩阵加上一个秩 1Hermite 矩阵或者通过加边造出一个新的 Hermite 矩阵，那么新旧矩阵的特征值必定交错. 关于它们的逆命题又怎样呢？如果给定两组交错的实数，能否造一个 Hermite 矩阵，以及一个对它作了适当修改的矩阵使得它们分别以这两组交错实数为其特征值呢？回答是肯定的. 作为例子. 我们给出定理(4.3.8)的逆定理.

4.3.10 定理 设 n 是给定的正整数，且设

$$\{\lambda_i : i=1,2,\cdots,n\}\quad 和\quad \{\hat{\lambda}_i : i=1,2,\cdots,n,n+1\}$$

[186]

是适合关系

$$\hat{\lambda}_1 \leqslant \lambda_1 \leqslant \hat{\lambda}_2 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_{n-1} \leqslant \hat{\lambda}_n \leqslant \lambda_n \leqslant \hat{\lambda}_{n+1}$$

的两个给定的实数序列. 设 $\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. 则存在实数 a 和实向量 $y\in\mathbf{R}^n$，使得 $\{\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_{n+1}\}$ 是实对称矩阵

$$\hat{A}=\begin{bmatrix}\Lambda & y\\ y^T & a\end{bmatrix}\in M_{n+1}(\mathbf{R})$$

的特征值集合.

证明：显然 $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ 是 Λ 的特征值集合，又因为 $\mathrm{tr}\,\hat{A}=\mathrm{tr}\,\Lambda+a$，我们必定有 $a=\mathrm{tr}\,\hat{A}-\mathrm{tr}\,\Lambda=\sum_{i=1}^{n+1}\hat{\lambda}_i-\sum_{i=1}^{n}\lambda_i$. 计算 $\hat{A}$ 的特征多项式 $p_{\hat{A}}(t)$ 是容易的，这是因为

$$\begin{aligned}\det(tI-\hat{A}) &= \det\begin{bmatrix}tI-\Lambda & -y\\ -y^T & t-a\end{bmatrix}\\
&=\det\begin{bmatrix}I & 0\\ [(tI-\Lambda)^{-1}y]^T & 1\end{bmatrix}\begin{bmatrix}tI-\Lambda & -y\\ -y^T & t-a\end{bmatrix}\begin{bmatrix}I & (tI-\Lambda)^{-1}y\\ 0 & 1\end{bmatrix}\\
&=\det\begin{bmatrix}tI-\Lambda & 0\\ 0 & (t-a)-y^T(tI-\Lambda)^{-1}y\end{bmatrix}\\
&=[(t-a)-y^T(tI-\Lambda)^{-1}y]\det(tI-\Lambda)\\
&=\left[(t-a)-\sum_{i=1}^{n}y_i^2\frac{1}{t-\lambda_i}\right]\prod_{i=1}^{n}(t-\lambda_i)=p_{\hat{A}}(t).\end{aligned} \tag{4.3.11}$$

我们已经确定了值 a，因此，剩下要证明的是，n 个实数 y_i 可以由(4.3.11)来确定，使得对 $k=1, 2, \cdots, n+1$ 有 $p_{\hat{A}}(\hat{\lambda}_k)=0$.

定义多项式

$$f(t)=\prod_{i=1}^{n+1}(t-\hat{\lambda}_i),\quad f\text{ 的次数}=n+1, \tag{4.3.12}$$

$$g(t)=\prod_{i=1}^{n}(t-\lambda_i),\quad g\text{ 的次数}=n. \tag{4.3.13}$$

根据 Euclid 算法，一定有

$$f(t)=g(t)(t-c)+r(t),$$

其中，c 是实数，而 $r(t)$ 是次数至多为 $n-1$ 的多项式，通过直接计算，求得 $c=\sum_{i=1}^{n+1}\hat{\lambda}_i-\sum_{i=1}^{n}\lambda_i=a$. 另外，因为 $g(\lambda_k)=0$，所以对于 $k=1, 2, \cdots, n$，$f(\lambda_k)=g(\lambda_k)(\lambda_k-a)+r(\lambda_k)=r(\lambda_k)$. 因此，就知道了多项式 $r(t)$ 在 n 个点的值，如果插值点 $\lambda_1, \cdots, \lambda_n$ 是互不相同的，则 $r(t)$ 可以用 Lagrange 插值公式明确写出. 在这个假定下，$g(t)$ 只有单根，因而，关于 $r(t)$ 的 Lagrange 插值公式是

187

$$r(t)=\sum_{i=1}^{n}f(\lambda_i)\frac{g(t)}{g'(\lambda_i)(t-\lambda_i)}.$$

于是，

$$\frac{f(t)}{g(t)}=(t-a)+\frac{r(t)}{g(t)}=(t-a)-\sum_{i=1}^{n}\frac{-f(\lambda_i)}{g'(\lambda_i)}\frac{1}{t-\lambda_i}$$

因为，对所有 $k=1, 2, \cdots, n+1$，$f(\hat{\lambda}_k)=0$，就必定有

$$(\hat{\lambda}_k - a) - \sum_{i=1}^{n} \frac{-f(\lambda_i)}{g'(\lambda_i)} \frac{1}{\hat{\lambda}_k - \lambda_i} = 0, \quad k=1,2,\cdots,n+1. \tag{4.3.14}$$

我们看到，如果对 $i=k-1$ 或 k 有 $\hat{\lambda}_k=\lambda_i$，则相应的项 $1/(t-\lambda_i)$ 有零系数，因而在 $t=\hat{\lambda}_k$ 时不会有奇异性. 如果对 $i=1, \cdots, n$，可以命 $y_i^2 \equiv -f(\lambda_i)/g'(\lambda_i)$，那么(4.3.11)保证 $p_{\hat{A}}(\hat{\lambda}_k)=0$，于是就完成了证明. 因此我们必须证明，$f(\lambda_i)/g'(\lambda_i) \leqslant 0$，$i=1, 2, \cdots, n$. 而现在该用交错性假设了. 利用 $f(t)$ 和 $g(t)$ 的定义及交错假设，得知

$$f(\lambda_i) = (-1)^{n-i+1} \prod_{j=1}^{n+1} |\lambda_i - \hat{\lambda}_j|,$$

$$g'(\lambda_i) = (-1)^{n-i} \prod_{\substack{j=1 \\ j\neq i}}^{n} |\lambda_i - \lambda_j|,$$

因而 $f(\lambda_i)$ 和 $g'(\lambda_i)$ 总有相反的符号.

不难修改上述论证以适用于有些 λ_i 项相等的情形. 例如，如果对某个 $k\geqslant 2$，有 $\lambda_1=\lambda_2=\cdots=\lambda_k<\lambda_{k+1}\leqslant\cdots$，则 $\hat{\lambda}_2=\cdots=\hat{\lambda}_k=\lambda_1$. (4.3.12)中的多项式 $g(t)$ 有因子 $(t-\hat{\lambda}_1)(t-\lambda_1)^{k-1}$；(4.3.13)中的多项式 $g(t)$ 有因子 $(t-\lambda_1)^k$，且 k 恰好是 λ_1 作为 $f(t)$ 的零点的重数. 因此，可以用 $(t-\lambda_1)^{k-1}$ 除 $f(t)$，$g(t)$ 和 $r(t)$ 来修改这每一个多项式. 修改后的多项式 $g(t)$ 将以 λ_1 作为其单重零点. 如果继续用这种办法取消 $g(t)$ 的所有重根，则证明可以如前继续下去，且结论是相同的. □

上述各个结果讨论了在一个 Hermite 矩阵旁"镶"上新的最后一行和最后一列的情形，但是这也可看成是给出了当把一个 Hermite 矩阵的最后一行和最后一列划去后其特征值的变化信 [188] 息，当然最后一行或一列并没什么特殊的. 在(4.3.8)中，如果划去矩阵 $\hat{A}$ 的第 i 行和第 i 列，而不是第 $(n+1)$ 行和 $(n+1)$ 列，那么在证明中，只是使 e_{n+1} 变成 e_i，并且得到相同的交错不等式组(4.3.9).

把定理(4.3.8)与(4.3.10)合在一起就说明，交错不等式组(4.3.9)是 Hermite 矩阵的诸特征值与它的任一 $n-1$ 阶主子矩阵的诸特征值之间的相互关系的一个完整的描述. 如果同时考虑 A 的所有 n 个 $(n-1)(n-1)$ 主子矩阵，那么还可以多说几句. 设 A_j 表示划去 A 的第 j 行和第 j 列后得到的主子矩阵，$j=1, 2, \cdots, n$，且设 A 和 A_j 的诸特征值按递增顺序排列. 对于每个 $i=1, 2, \cdots, n-1$，有

$$\max_{1\leqslant j\leqslant n} \lambda_i(A_j) \geqslant \frac{n-i}{n}\lambda_i(A) + \frac{i}{n}\lambda_{i+1}(A),$$

$$\min_{1\leqslant j\leqslant n} \lambda_i(A_j) \leqslant \frac{n-i}{n}\lambda_i(A) + \frac{i}{n}\lambda_n(A),$$

和

$$\max_{1\leqslant j\leqslant n} \lambda_{n-1}(A_j) - \min_{1\leqslant j\leqslant n} \lambda_1(A_j) \geqslant \left(\frac{n-2}{n}\right)^{1/2}[\lambda_n(A) - \lambda_1(A)].$$

如果 A 的所有特征值非负，即如果 A 是半正定矩阵，那么由这三个不等式中的第一个推出，至少有一个主子矩阵 A_j，可使

$$\lambda_{n-1}(A_i) \geqslant \frac{n-1}{n}\lambda_n(A).$$

因此，半正定 Hermite 矩阵的每个主子矩阵的谱半径不会"很小".

人们可能希望从一个 Hermite 矩阵中划去若干行和若干相应的列，余下的矩阵是原矩阵的一个主子矩阵. 下述结果可以通过重复应用交错不等式组(4.3.9)来得到，但是从 Courant-Fischer 定理直接证明这个论断也一样容易. 有时称这个结果为*包含原理*.

4.3.15 定理 设 $A\in M_n$ 是 Hermite 矩阵，r 是整数，$1\leqslant r\leqslant n$，又设 A_r 表示(从 A 中划去 $n-r$ 行和相应的 $n-r$ 列后得到的)A 的任一 $r\times r$ 主子矩阵. 则对于使 $1\leqslant k\leqslant r$ 的每个整数 k，有

189

$$\lambda_k(A) \leqslant \lambda_k(A_r) \leqslant \lambda_{k+n-r}(A).$$

证明： 假定 $A_r\in M_r$ 是从 A 中划去 i_1、…，i_{n-r} 行和相应的诸列后得到的，且设 $1\leqslant k\leqslant r$. 利用(4.2.12)可导出

$$\begin{aligned}\lambda_{k+n-r}(A) &= \min_{w_1,\cdots,w_{r-k}\in \mathbf{C}^n}\ \max_{\substack{x\neq 0,x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{r-k}}} \frac{x^*Ax}{x^*x}\\ &\geqslant \min_{w_1,\cdots,w_{r-k}\in \mathbf{C}^n}\ \max_{\substack{x\neq 0,x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{r-k}\\ x\perp e_{i_1},\cdots,e_{i_{n-r}}}} \frac{x^*Ax}{x^*x}\\ &= \min_{v_1,\cdots,v_{r-k}\in \mathbf{C}^r}\ \max_{\substack{y\neq 0,y\in\mathbf{C}^r\\ y\perp v_1,\cdots,v_{r-k}}} \frac{y^*A_ry}{y^*y} = \lambda_k(A_r).\end{aligned}$$

假定 $1\leqslant k\leqslant r$，再利用(4.2.13)可导出

$$\begin{aligned}\lambda_k(A) &= \max_{w_1,\cdots,w_{k-1}\in \mathbf{C}^n}\ \min_{\substack{x\neq 0,x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{k-1}}} \frac{x^*Ax}{x^*x}\\ &\leqslant \max_{w_1,\cdots,w_{k-1}\in \mathbf{C}^n}\ \min_{\substack{x\neq 0,x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{k-1}\\ x\perp e_{i_1},\cdots,e_{i_{n-r}}}} \frac{x^*Ax}{x^*x}\\ &= \max_{v_1,\cdots,v_{k-1}\in \mathbf{C}^r}\ \min_{\substack{y\neq 0,y\in\mathbf{C}^r\\ y\perp v_1,\cdots,v_{k-1}}} \frac{y^*A_ry}{y^*y} = \lambda_k(A_r).\end{aligned}$$

□

定理(4.3.15)的下述简单推论称为 Poincaré *分离定理*. 它可以应用于(像量子力学)那样的场合，在那里人们有关于内积 $u_i^*Au_j$ 方面的信息.

4.3.16 推论 设 $A\in M_n$ 是 Hermite 矩阵，r 是适合 $1\leqslant r\leqslant n$ 的某个整数，设 u_1，…，$u_r\in\mathbf{C}^n$ 是 r 个给定的单位正交向量. 设 $B_r=[u_i^*Au_j]\in M_r$. 如果 A 和 B_r 的诸特征值按递增顺序(4.2.1)排列，则有

$$\lambda_k(A)\leqslant\lambda_k(B_r)\leqslant\lambda_{k+n-r}(A),\quad k=1,2,\cdots,r. \tag{4.3.17}$$

证明： 如果 $r<n$，另选 $n-r$ 个向量 u_{r+1}，…，u_n 使得$\{u_1$，…，u_r，u_{r+1}，…，$u_n\}$是标准正交组，设 $U=[u_1$，…，$u_n]\in M_n$. 矩阵 U 是酉矩阵，U^*AU 与 A 有相同的特征值，而已知矩阵 B_r 是划去最后 $n-r$ 行和列后得到的 U^*AU 的主子矩阵. 现在，论断可从(4.3.15)得

190

出. □

在上述结果中，矩阵 $B_r \in M_r$ 可以写成 $U^* AU$，其中 $U \in M_{n,r}$ 是有 r 个单位正交列的矩阵，因为 $\operatorname{tr} B_r = \lambda_1(B_r) + \cdots + \lambda_r(B_r)$，所以，下述极值特征可以通过相加不等式(4.3.17)来得到.

4.3.18 推论 设 $A \in M_n$ 是 Hermite 矩阵，r 是适合 $1 \leqslant r \leqslant n$ 的某个整数，那么

$$\lambda_1(A) + \cdots + \lambda_r(A) = \min_{U^* U = I \in M_r} \operatorname{tr} U^* AU, \quad U \in M_{n,r}. \tag{4.3.19}$$

$$\lambda_{n-r+1}(A) + \cdots + \lambda_n(A) = \max_{U^* U = I \in M_r} \operatorname{tr} U^* AU, \quad U \in M_{n,r}. \tag{4.3.20}$$

如果 U 的诸列选为 A 的 r 个最小特征值的相应单位正交向量，则(4.3.19)中等式成立. 类似的选择推出(4.3.20)的等式成立. 这两个不等式可以看作 Rayleigh-Ritz 定理(4.2.2)的推广，它们可以用来证明许多其他有意义的不等式.

有时，我们知道二次型 $x^* Ax$ 在一个子空间上的变化范围，把 Courant-Fischer 定理用于这种情形便可以给出 A 的诸特征值的界.

4.3.21 定理 设 $A \in M_n$ 是 Hermite 矩阵，k 是适合 $1 \leqslant k \leqslant n$ 的某个整数. 设 A 的诸特征值按递增顺序(4.2.1)排列，且设 S_k 是 $\mathbf{C}^n$ 的某个 k 维子空间. 如果存在常数 c_2，使得对所有 $x \in S_k$ 有 $x^* Ax \geqslant c_2 x^* x$，则 $\lambda_n \geqslant \lambda_{n-1} \geqslant \cdots \geqslant \lambda_{n-k+1} \geqslant c_2$. 如果存在常数 c_1，使得对所有 $x \in S_k$ 有 $x^* Ax \leqslant c_1 x^* x$，则 $c_1 \geqslant \lambda_k \geqslant \cdots \geqslant \lambda_1$.

证明：设 $u_1, \cdots, u_{n-k}$ 是张成 $S_k^{\perp}$ 的 $n-k$ 个单位正交向量. 利用(4.2.13)可导出

$$c_2 \leqslant \min_{\substack{x \neq 0 \\ x \in S_k}} \frac{x^* Ax}{x^* x} = \min_{\substack{x \neq 0 \\ x \perp u_1, \cdots, u_{n-k}}} \frac{x^* Ax}{x^* x}$$

$$\leqslant \max_{w_1, \cdots, w_{n-k} \in \mathbf{C}^n} \min_{\substack{x \neq 0 \\ x \perp w_1, \cdots, w_{n-k}}} \frac{x^* Ax}{x^* x} = \lambda_{n-k+1}. \tag{4.3.22}$$

类似地，可以利用(4.2.12)导出

191

$$c_1 \geqslant \max_{\substack{x \neq 0 \\ x \in S_k}} \frac{x^* Ax}{x^* x} = \max_{\substack{x \neq 0 \\ x \perp u_1, \cdots, u_{n-k}}} \frac{x^* Ax}{x^* x}$$

$$\geqslant \min_{w_1, \cdots, w_{n-k} \in \mathbf{C}^n} \max_{\substack{x \neq 0 \\ x \perp w_1, \cdots, w_{n-k}}} \frac{x^* Ax}{x^* x} = \lambda_k.$$ □

4.3.23 推论 如果 $A \in M_n$ 是 Hermite 矩阵，且对 k 维子空间的所有向量 x 有 $x^* Ax \geqslant 0$，则 A 至少有 k 个非负特征值. 如果对 k 维子空间的所有非零向量有 $x^* Ax > 0$，则 A 至少有 k 个正特征值.

证明：第一个论断可由上述定理推出，只需取 $c_2 = 0$. 如果 $\lambda_{n-k+1} = 0$，则不等式(4.3.22)说明

$$0 = \min_{\substack{x \neq 0 \\ x \in S_k}} \frac{x^* Ax}{x^* x} = \min_{\substack{x^* x = 1 \\ x \in S_k}} x^* Ax.$$

但 S_k 是有限维的，所以集 $D = \{x \in S_k: x^* x = 1\}$ 是紧集[见(5.5.6)]，因而连续函数 $x^* Ax$ 对

某个满足 $x^*x=1$ 的 $x_0\in S_k$ 达到它在 D 上的极小值；特别是 $x_0\neq 0$．然而，$x_0^*Ax_0=0$，这与当 $x\in S_k$ 且 $x\neq 0$ 时有 $x^*Ax>0$ 的假设相矛盾． □

Hermite 矩阵的诸特征值和诸主对角元都是实数，且诸特征值之和与诸对角元之和(迹)相同．诸主对角元和诸特征值间的明确关系可用优化概念给出．

4.3.24 定义 设 $\alpha=[\alpha_i]\in\mathbf{R}^n$ 和 $\beta=[\beta_i]\in\mathbf{R}^n$ 已知．如果对所有 $k=1,2,\cdots,n$，

$$\min\left\{\sum_{j=1}^{k}\beta_{i_j}:1\leqslant i_1<\cdots<i_k\leqslant n\right\}\geqslant\min\left\{\sum_{j=1}^{k}\alpha_{i_j}:1\leqslant i_1<\cdots<i_k\leqslant n\right\},$$

且当 $k=n$ 时等式成立，则称向量 β 优化向量 α．如果按递增顺序 $\alpha_{j_1}\leqslant\alpha_{j_2}\leqslant\cdots\leqslant\alpha_{j_n}$，$\beta_{m_1}\leqslant\beta_{m_2}\leqslant\cdots\leqslant\beta_{m_n}$ 排列 α 和 β 的各元素，则定义的不等式组可以表达成等价的形式

$$\sum_{i=1}^{k}\beta_{m_i}\geqslant\sum_{i=1}^{k}\alpha_{j_i}.\tag{4.3.25}$$

192 它对所有 $k=1,2,\cdots,n$ 成立，且等式对 $k=n$ 成立．

因此，实向量 β 优化实向量 α，是指对 $k=1,2,\cdots,n-1$，β 的 k 个最小元素之和大于或等于 α 的 k 个最小元素之和，且 β 与 α 的各元素之和相等．应指出的是，可以任意改变 β 与 α 各元素的顺序，而不影响 β 是否优化 α．

优化概念是在矩阵理论的许多场合中作为两个实数集间的确定关系产生出来的重要概念．这种情形的一个例子是下述 Schur 定理(1923)．

4.3.26 定理 设 $A\in M_n$ 是 Hermite 矩阵，则 A 的诸对角元组成的向量优化 A 的诸特征值组成的向量．

证明： 证明是对维数作归纳法．对 $n=1$，没有什么可证的，所以，假定结论对所有 $k\leqslant n-1$ 维 Hermite 矩阵成立．设 $A=[a_{ij}]\in M_n$ 是给定的 Hermite 矩阵，且设 $A_1\in M_{n-1}$ 是划去对应于 A 的最大对角元的行和列后得到的 A 的主子矩阵．设 $\lambda_1\leqslant\cdots\leqslant\lambda_n$ 是 A 的有序特征值，$\lambda_1'\leqslant\cdots\leqslant\lambda_{n-1}'$ 是 A_1 的有序特征值，且设 $a_{i_1i_1}\leqslant a_{i_2i_2}\leqslant\cdots\leqslant a_{i_ni_n}$ 是诸对角元按递增顺序的重排．根据归纳假定，对所有 $k=1,\cdots,n-1$，有

$$\sum_{j=1}^{k}a_{i_ji_j}\geqslant\sum_{j=1}^{k}\lambda_j'.$$

因为交错性[定理(4.3.8)]，还有

$$\lambda_1\leqslant\lambda_1'\leqslant\lambda_2\leqslant\lambda_2'\leqslant\cdots\leqslant\lambda_{n-1}'\leqslant\lambda_n,$$

因而，对所有 $k=1,2,\cdots,n-1$，有

$$\sum_{j=1}^{k}\lambda_j'\geqslant\sum_{j=1}^{k}\lambda_j.$$

因此

$$\sum_{j=1}^{k}a_{i_ji_j}\geqslant\sum_{j=1}^{k}\lambda_j,\quad k=1,\cdots,n-1,$$

又因为迹是所有特征值的和，所以等式对 $k=n$ 成立． □

193 优化概念也可以用来表示矩阵之和的诸特征值与其被加矩阵的诸特征值间的关系．

4.3.27 定理 设 A，$B\in M_n$ 是 Hermite 矩阵，设 $\lambda(A)=[\lambda_i(A)]$，$\lambda(B)=[\lambda_i(B)]$和 $\lambda(A+B)=[\lambda_i(A+B)]$表示 $\mathbf{R}^n$ 中的列向量，它们的诸分量各是按递增顺序(4.2.1)排列的 A，B 和 $A+B$ 的诸特征值，则向量 $\lambda(A+B)$优化向量 $\lambda(A)+\lambda(B)$.

证明：对任意 $k=1$，2，…，n，利用(4.3.18)可导出

$$\begin{aligned}\sum_{i=1}^{k}\lambda_i(A+B)&=\min_{U^*U=I\in M_k}\operatorname{tr}U^*(A+B)U\\&=\min_{U^*U=I\in M_k}(\operatorname{tr}U^*AU+\operatorname{tr}U^*BU)\\&\geqslant\min_{U^*U=I\in M_k}\operatorname{tr}U^*AU+\min_{U^*U=I\in M_k}\operatorname{tr}U^*BU\\&=\sum_{i=1}^{k}\lambda_i(A)+\sum_{i=1}^{k}\lambda_i(B)=\sum_{i=1}^{k}(\lambda_i(A)+\lambda_i(B)).\end{aligned}$$

因为 $\operatorname{tr}(A+B)=\operatorname{tr}A+\operatorname{tr}B$，对于 $k=n$，等式成立. □

我们已经说过，优化是 Hermite 矩阵的诸主对角元与它的诸特征值间的一个确定的关系，但是，在(4.3.26)中只是建立起这种关系的一半，为了证明另一半，需要下面的技巧性引理.

4.3.28 引理 设 $n\geqslant2$，且设 $\alpha_1\leqslant\alpha_2\leqslant\cdots\leqslant\alpha_n$ 和 $\beta_1\leqslant\beta_2\leqslant\cdots\leqslant\beta_n$ 是给定的实数. 如果向量 $\beta=[\beta_i]$优化向量 $\alpha=[\alpha_i]$，则存在 $n-1$ 个实数 γ_1，…，γ_{n-1}，使得

$$\alpha_1\leqslant\gamma_1\leqslant\alpha_2\leqslant\gamma_2\leqslant\alpha_3\leqslant\cdots\leqslant\alpha_{n-1}\leqslant\gamma_{n-1}\leqslant\alpha_n,$$

且使得 $\beta'=[\beta_1,\cdots,\beta_{n-1}]^T\in\mathbf{R}^{n-1}$优化 $\gamma=[\gamma_1,\cdots,\gamma_{n-1}]^T\in\mathbf{R}^{n-1}$.

证明：如果 $n=2$，则有 $\alpha_1\leqslant\beta_1$ 且 $\alpha_1+\alpha_2=\beta_1+\beta_2$，或

$$\alpha_2=(\beta_1-\alpha_1)+\beta_2\geqslant\beta_2\geqslant\beta_1.$$

所以，$\alpha_1\leqslant\beta_1\leqslant\alpha_2$，因而可以选取 $\gamma_1=\beta_1$ 且适合所述条件. 现在假定 $n\geqslant2$，且设 $\Delta=\{[\delta_1,\cdots,\delta_{n-1}]^T\}\subset\mathbf{R}^{n-1}$表示用不等式组

$$\alpha_1\leqslant\delta_1\leqslant\alpha_2\leqslant\delta_2\leqslant\alpha_3\leqslant\cdots\leqslant\alpha_{n-1}\leqslant\delta_{n-1}\leqslant\alpha_n,\tag{4.3.29a}$$

$$\sum_{i=1}^{k}\delta_i\leqslant\sum_{i=1}^{k}\beta_i,\quad k=1,2,\cdots,n-2\tag{4.3.29b}$$

194

定义的点集. 因为 β 优化 α，所以点 $\delta=\hat{\alpha}=[\alpha_1,\cdots,\alpha_{n-1}]^T$ 总在 Δ 中，因而集 Δ 一定不空. Δ 显然是有界闭集，且容易看出它是凸集. 如果 $\delta=[\delta_1,\cdots,\delta_{n-1}]^T\in\Delta$，则定义 $f(\delta)\equiv\delta_1+\delta_2+\cdots+\delta_{n-1}$. 注意，$f(\hat{\alpha})=\alpha_1+\cdots+\alpha_{n-1}\leqslant\beta_1+\cdots+\beta_{n-1}$. 如果能够证明有某个 $\hat{\delta}\in\Delta$ 使 $f(\hat{\delta})\geqslant\beta_1+\cdots+\beta_{n-1}$，则根据 Δ 的凸性，对所有 $t\in[0,1]$，将有 $t\hat{\alpha}+(1-t)\hat{\delta}\in\Delta$，并且 $g(t)\equiv f(t\hat{\alpha}+(1-t)\hat{\delta})$将是适合

$$g(0)\geqslant\beta_1+\cdots+\beta_{n-1}\geqslant g(1)$$

的连续函数. 由此可以得出，对某个 $t_0\in[0,1]$，有 $g(t_0)=\beta_1+\cdots+\beta_{n-1}$. 点 $\gamma=[\gamma_i]=t_0\hat{\alpha}+(1-t_0)\hat{\delta}$ 将适合所述条件.

因为 $f(\cdot)$是紧集 Δ 上的连续函数，所以存在点 $\hat{\delta}\in\Delta$，使得

$$\max_{\delta\in\Delta}f(\delta)=f(\hat{\delta}).\tag{4.3.30}$$

我们要证明 $f(\hat{\delta})\geqslant\beta_1+\cdots+\beta_{n-1}$. 极大值点 $\hat{\delta}\in\Delta$ 适合不等式组(4.3.29a 和 b)，因而

$$\hat{\delta}_k \leqslant \alpha_{k+1}, \quad k = 1,2,\cdots,n-1, \tag{4.3.31a}$$

$$\sum_{i=1}^{k} \hat{\delta}_i \leqslant \sum_{i=1}^{k} \beta_i, \quad k = 1,2,\cdots,n-2. \tag{4.3.31b}$$

如果整个不等式组(4.3.31b)是严格的，且不等式组(4.3.31a)中至少有一个不是等式，则 $\hat{\delta}$ 至少有一个分量还可以增大以使 $f(\hat{\delta})$ 的值相应增大. 因为这与极值性质(4.3.30)相矛盾，所以得出所有不等式(4.3.31a)必须都是等式. $\hat{\delta}=[\alpha_2, \alpha_3, \cdots, \alpha_n]^T$，因而 $f(\hat{\delta})=\alpha_2+\cdots+\alpha_n=(\alpha_1+\alpha_2+\cdots+\alpha_n)-\alpha_1=(\beta_1+\beta_2+\cdots+\beta_n)-\alpha_1=(\beta_1+\cdots+\beta_{n-1})+\beta_n-\alpha_1\geqslant(\beta_1+\cdots+\beta_{n-1})+\beta_1-\alpha_1\geqslant\beta_1+\cdots+\beta_{n-1}$，这正是我们想要证明的.

如果不等式组(4.3.31b)不都是严格的，则至少有 k 的一个值使等式成立. 设 r 表示这种 k 值中最大的一个，于是

$$\sum_{i=1}^{r} \hat{\delta}_i = \sum_{i=1}^{r} \beta_i,$$

$$\sum_{i=1}^{k} \hat{\delta}_i < \sum_{i=1}^{k} \beta_i, \quad k = r+1,\cdots,n-2.$$

195 根据在上一段中相同的论证，对 $k=r+1, \cdots, n-1$，必定有 $\hat{\delta}_k=\alpha_{k+1}$. 因此，

$$\begin{aligned}
f(\hat{\delta}) &= (\hat{\delta}_1+\cdots+\hat{\delta}_r)+(\hat{\delta}_{r+1}+\cdots+\hat{\delta}_{n-1}) \\
&= (\beta_1+\cdots+\beta_r)+(\alpha_{r+2}+\cdots+\alpha_n) \\
&= (\beta_1+\cdots+\beta_{n-1})+(\alpha_1+\cdots+\alpha_n)-(\alpha_1+\cdots+\alpha_{r+1})-(\beta_{r+1}+\cdots+\beta_{n-1}) \\
&= (\beta_1+\cdots+\beta_{n-1})+(\beta_1+\cdots+\beta_n)-(\alpha_1+\cdots+\alpha_{r+1})-(\beta_{r+1}+\cdots+\beta_{n-1}) \\
&= (\beta_1+\cdots+\beta_{n-1})+[(\beta_1+\cdots+\beta_{r+1})-(\alpha_1+\cdots+\alpha_{r+1})]+(\beta_{r+2}+\cdots+\beta_n) \\
&\quad -(\beta_{r+1}+\cdots+\beta_{n-1}) \\
&\geqslant (\beta_1+\cdots+\beta_{n-1})+(\beta_{r+2}-\beta_{r+1})+(\beta_{r+3}-\beta_{r+2})+\cdots+(\beta_n-\beta_{n-1}) \\
&\geqslant \beta_1+\cdots+\beta_{n-1}.
\end{aligned}$$

□

我们现在可以证明(4.3.26)的逆命题了.

4.3.32 定理 设 $n\geqslant1$，且设 $a_1\leqslant a_2\leqslant\cdots\leqslant a_n$ 和 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$ 是给定的实数，如果向量 $a=[a_i]$ 优化向量 $\lambda=[\lambda_i]$，则有实对称矩阵 $A=[a_{ij}]\in M_n(\mathbf{R})$，使得 $a_{ii}=a_i$ 对 $i=1, 2, \cdots, n$ 成立，且使得 $\{\lambda_i\}$ 是 A 的特征值集合.

证明： 对 $n=1$，论断显然成立，假设对至多有 $n-1$ 个元素的所有这些向量 a 和 λ，论断已经证明. 根据引理，存在实数 $\gamma_1\leqslant\gamma_2\leqslant\cdots\leqslant\gamma_{n-1}$，使得

$$\lambda_1 \leqslant \gamma_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_{n-1} \leqslant \gamma_{n-1} \leqslant \lambda_n,$$

且使 $a'=[a_1, \cdots, a_{n-1}]^T$ 优化 $\gamma=[\gamma_i]^T\in\mathbf{R}^{n-1}$. 根据归纳假定，存在实对称矩阵 $B=[b_{ij}]\in M_{n-1}$，使得 $\{\gamma_i\}$ 是 B 的特征值集合，且对 $i=1, 2, \cdots, n-1$ 有 $b_{ii}=a_i$. 如果 $\Gamma=\mathrm{diag}(\gamma_1, \gamma_2, \cdots, \gamma_{n-1})\in M_{n-1}(\mathbf{R})$，则存在实正交矩阵 $Q\in M_{n-1}(\mathbf{R})$ 使得 $B=Q\Gamma Q^T$. 根据定理(4.3.10)，存在实对称矩阵

$$\hat{A}=\begin{bmatrix} \Gamma & y \\ y^T & \alpha \end{bmatrix}\in M_n(\mathbf{R}), \quad y\in\mathbf{R}^{n-1}, \quad \alpha\in\mathbf{R},$$

它有特征值$\{\lambda_i\}$. 如果令

$$A \equiv \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} \hat{A} \begin{bmatrix} Q^T & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} Q\Gamma Q^T & Qy \\ (Qy)^T & \alpha \end{bmatrix} = \begin{bmatrix} B & Qy \\ (Qy)^T & \alpha \end{bmatrix},$$

196

则 A 有特征值$\{\lambda_i\}$，且有主对角元 a_1，a_2，…，a_{n-1}，α. 但是，根据优化条件，$\mathrm{tr}\, A = a_1 + \cdots + a_{n-1} + \alpha = \lambda_1 + \cdots + \lambda_n = a_1 + \cdots + a_n$，因而 $\alpha = a_n$ 且 A 有符合要求的对角元. □

上述结果不仅完成了涉及 Hermite 矩阵的诸主对角元和诸特征值间的关系的推理过程，而且能够解释优化关系自身的几何意义. 双随机矩阵 $A \in M_n$ 有 n^2 个非负元，且每行和每列的诸元之和是 +1. Brikhoff 定理(8.7.1)保证每个双随机矩阵是有限多个置换矩阵的一个凸组合，反之亦然.

4.3.33 定理 设 $\alpha = [\alpha_i] \in \mathbf{R}^n$ 和 $\beta = [\beta_i] \in \mathbf{R}^n$ 是两个给定的实向量，则下列条件等价：

(a) β 优化 α；

(b) 存在双随机矩阵 $S \in M_n$ 使得 $\beta = S\alpha$；

(c) $\beta \in \{\sum_{i=1}^{N} p_i \alpha_{\pi_i}\}$，其中 $1 \leqslant N < \infty$，$p_i \geqslant 0$，$\sum_{i=1}^{N} p_i = 1$，而 $\alpha_{\pi_i} \in \mathbf{R}^n$ 是一个向量，它的诸分量是已知向量 α 的诸分量的某个置换.

证明： 如果假定(a)成立，则根据(4.3.32)，存在有主对角元 $b_{ii} = \beta_i$ 和特征值 $\lambda_i(B) = \alpha_i$ 的实对称矩阵. 根据谱定理，有酉(甚至实正交)矩阵 $U = [u_{ij}] \in M_n$ 使得 $B = U\Lambda U^*$，其中 $\Lambda = \mathrm{diag}(\alpha_1, \cdots, \alpha_n)$，而考察 B 的诸主对角元可知 $\beta = S\alpha$，其中 $S = [s_{ij}] \in M_n$ 是由 $S_{ij} = |u_{ij}|^2$ 给定的. 因为 U 的每行和每列是单位向量，这个 S 的每个行和及每个列和都等于 1，所以 S 是双随机矩阵(称这个特殊形式的矩阵为正交随机矩阵)，这就证明了(a)蕴涵(b).

(b) 蕴涵(a)的证明在本节末的习题 9 中概述.

如果假定(b)成立，则 Birkhoff 定理(8.7.1)说明，

$$S = \sum_{i=1}^{N} p_i P_i, \quad 其中\ p_i \geqslant 0, \quad \sum_{i=1}^{N} p_i = 1,$$

而每个 P_i 是置换矩阵. 因此，

$$\beta = S\alpha = \sum_{i=1}^{N} p_i P_i \alpha = \sum_{i=1}^{N} p_i \alpha_{\pi_i}, \quad 其中\ P_i \alpha \equiv \alpha_{\pi_i}.$$

这个恒等式也证明了逆蕴涵关系. □

197

这样一来，由某个向量 $\beta = [\beta_1, \cdots, \beta_n]^T$ 优化的所有向量 $\alpha = [\alpha_1, \cdots, \alpha_n]^T$ 组成的集合，可以先计算出 $n!$ 个向量然后作这些向量的凸包来得到，而这 $n!$ 个向量是采用置换 β 向量的 n 个分量的办法得来的(当然，如果诸项 β_i 不全不同，则这些向量也不全不同).

附注 虽然一致赞同优化这一基本概念是很重要的，但采用的记号常不一致. 有些作者用(4.3.25)中的反向不等式定义优化，而有些作者定义优化是按递减顺序排列两个集合. 因此，当人们从不同的文献中采用或引用有关优化的结果时，一定要倍加小心. 参看习题 11 中促使我们选择优化定义的理由.

习题

1. 我们知道，矩阵 $A\in M_n$ 的谱半径是数值

$$\rho(A)=\max\{|\lambda_i(A)|\}.$$

设 A，$B\in M_n$ 是 Hermite 矩阵. 试用 Weyl 定理(4.3.1)证明

$$\lambda_1(B)\leqslant\lambda_k(A+B)-\lambda_k(A)\leqslant\lambda_n(B),$$

因而，对所有 $k=1,2,\cdots,n$ 有

$$|\lambda_k(A+B)-\lambda_k(A)|\leqslant\rho(B).$$

这是关于 Hermite 矩阵的特征值的扰动定理[参看(6.3)]的一个简单例子.

2. 试说明仅利用定理(4.3.4)证明中所导出的第一组不等式如何得到定理中所确定的所有不等式. **提示：** $A=(A\pm zz^*)\mp zz^*$.

3. 详细证明(4.3.6b)等价于(4.3.6a).

4. Weyl 定理(4.3.7)的证明中只用到了不等式 $\lambda_n(A+B)\geqslant\lambda_{n-r}(A)$，其中 B 至多有秩 r. 证明这个不等式无须利用 Courant-Fischer 定理，而是用需要提供详细论述的下述论断. 假定 $B=\beta_1y_1y_1^*+\cdots+\beta_ry_ry_r^*$，并且设 $A=U\Lambda U^*$，其中 $U=[u_1\cdots u_n]$是酉矩阵. 则存在 $r+1$ 个纯量 α_{n-r}，α_{n-r+1}，…，α_n，使得对所有 $i=1,2,\cdots,r$，向量 $x=\alpha_{n-r}u_{n-r}+\cdots+\alpha_nu_n$ 适合 $x\perp y_i$，且 $x^*x=|\alpha_{n-r}|^2+\cdots+|\alpha_n|^2=1$. 于是

198

$$\lambda_n(A+B)\geqslant x^*(A+B)x=\sum_{i=n-r}^{n}|\alpha_i|^2\lambda_i(A)\geqslant\lambda_{n-r}(A).$$

5. 试应用定理(4.3.8)$n-r$ 次来证明定理(4.3.15).

6. 试说明，如果 A 和 B 不是 Hermite 矩阵，则 Weyl 的简单不等式组不一定成立. **提示：** 考虑 $A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$和 $B=\begin{bmatrix}0&0\\1&0\end{bmatrix}$.

7. 如果 A，$B\in M_n$ 是 Hermite 矩阵，它们的特征值按递增顺序排列，又如果 $1\leqslant k\leqslant n$，证明 $\lambda_k(A+B)\leqslant\min\{\lambda_i(A)+\lambda_j(B):\ i+j=k+n\}$.

8. 试对定理(4.3.10)的证明中有相同项 λ_i 的情形给出详细说明. **提示：** 如果 λ_1 是 $g(t)=0$ 的 k 重根，$k\geqslant2$，证明，在得到的(4.3.14)中，(4.3.14)的分母中的 $g'(t)$ 的因式 $(t-\lambda_1)^{k-1}$ 可消去(4.3.14)的分子中 $f(t)$ 的因式 $(t-\lambda_1)^{k-1}$.

9. 设 $S=[s_{ij}]\in M_n$ 是双随机矩阵(8.7)，$x\in\mathbf{R}^n$ 是实向量. 证明 Sx 优化 x. **提示：** 设 $y=Sx$. 只要对 $y_1\leqslant\cdots\leqslant y_n$ 与 $x_1\leqslant\cdots\leqslant x_n$ 的情形证明就可以了，因为如果不是这样，就可以对适当的置换矩阵 P 和 Q 来考虑 Py 和 Qx；而 P^TSQ^T 仍是双随机矩阵. 设 $w_j^{(k)}=\sum_{i=1}^{k}s_{ij}$，于是，$0\leqslant w_j^{(k)}\leqslant1$，且 $\sum_{j=1}^{n}w_j^{(k)}=k$. 证明

$$\sum_{i=1}^{k}(y_i-x_i)=\sum_{j=1}^{n}w_j^{(k)}x_j-\sum_{i=1}^{k}x_i+x_k\Big(k-\sum_{j=1}^{n}w_j^{(k)}\Big)$$

$$=\sum_{i=1}^{k}(1-w_i^{(k)})(x_k-x_i)+\sum_{j=k+1}^{n}w_j^{(k)}(x_j-x_k).$$

10. 用下述思路给出定理(4.3.26)的另一个证明：若 $A=[a_{ij}]\in M_n$ 是 Hermite 矩阵，则 $A=U\Lambda U^*$，其中 $U=[u_{ij}]\in M_n$ 是酉矩阵，$\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ 是实对角矩阵. 设 $a=[a_{11}, a_{22}, \cdots, a_{nn}]^T$ 是由 A 的主对角元组成的向量，且 $x=[\lambda_1, \lambda_2, \cdots, \lambda_n]^T$. 证明 $a=Px$，其中 $P=[p_{ij}]=[|u_{ij}|^2]$. 证明 P 是双随机矩阵，然后利用习题 9.

11. 设 $x=[x_1, \cdots, x_n]^T$ 和 $y=[y_1, \cdots, y_n]^T$ 是两个给定的非负实向量，且假定 y 优化 x. 证明 $y_1\cdots y_n\geqslant x_1\cdots x_n$. **提示**：利用(4.3.32)构造一个实对称矩阵 $A=[a_{ij}]\in M_n(\mathbf{R})$，其中主对角元 $a_{ii}\equiv y_i$，而特征值 $\lambda_i(A)=x_i$. 则 Hadamard 不等式(7.8.1)是指 $a_{11}\cdots a_{nn}\geqslant\det A=\lambda_1\cdots\lambda_n$.

附注 这就是促使我们选择优化定义(4.3.24)的理由. 如果有人选择与(4.3.24)中的不等式相反的方向，则所得结论与习题 11 中的不等式相反，即，如果在这个意义下 y“优化”x，则 y_i 的乘积小于 x_i 的乘积. 我们喜欢的定义形式是，乘积不等式的走向与优化的方向相同.

199

12. 设 $A\in M_n$ 是具有正特征值 $0<\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$ 的 Hermite 矩阵，又设 $1\leqslant r\leqslant n$ 是一个给定的整数. 用习题 11 证明

$$\lambda_1\lambda_2\cdots\lambda_r = \min(u_1^*Au_1)(u_2^*Au_2)\cdots(u_r^*Au_r),$$

其中极小取遍所有标准正交组 $\{u_1, u_2, \cdots, u_r\}\subset\mathbf{C}^n$. 说明为什么这个结果可以相继看作(4.3.19)的乘法模拟，Hadamard 不等式(7.8.1)的推广，联系 Rayleigh-Ritz 定理(4.2.2)与 Hadamard 不等式的不等式系列. **提示**：情形 $r=n$ 是(7.8.1). 情形 $r=1$ 是什么？若 $2\leqslant r\leqslant n$，用(4.3.19)证明，向量 $[u_1^*Au_1, u_2^*Au_2, \cdots, u_r^*Au_r]^T$ 优化向量 $[\mu_1, \cdots, \mu_r]^T$，其中，对所有 $i=1, 2, \cdots, r-1$，$\mu_i=\lambda_i$，而

$$\mu_r=(u_1^*Au_1+\cdots+u_{r-1}^*Au_{r-1})-(\lambda_1+\cdots+\lambda_{r-1})+u_r^*Au_r\geqslant u_r^*Au_r.$$

然后用习题 11 证明 $\lambda_1\cdots\lambda_{r-1}u_r^*Au_r\leqslant\prod_{i=1}^{r}u_i^*Au_i$.

13. 设 $A=[a_{ij}]\in M_n$ 是具有非负特征值 $0\leqslant\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$ 的 Hermite 矩阵. 证明，对每个 $r=1, 2, \cdots, n$，乘积 $\lambda_1\cdots\lambda_r$ 小于或等于 A 的 r 个最小主对角元之乘积.

14. 如果 A, $B\in M_n$ 是 Hermite 矩阵，$A-B$ 具有唯一非负特征值，证明对所有 $i=1, 2, \cdots, n$ 有 $\lambda_i(A)\geqslant\lambda_i(B)$.

15. 试用(4.3.18)证明(4.3.26). **提示**：用置换矩阵调整 $A=[a_{ij}]$ 使 $a_{11}\leqslant a_{22}\leqslant\cdots\leqslant a_{nn}$. 然后取 $U=[e_1e_2\cdots e_r]\in M_{n,r}$，并且证明 $\lambda_1(A)+\cdots+\lambda_r(A)\leqslant\mathrm{tr}\,U^*AU=a_{11}+\cdots+a_{rr}$.

16. 假定 $A\in M_n$ 是 Hermite 矩阵，设 $\lambda_1\leqslant\cdots\leqslant\lambda_n$ 是 A 的特征值，而 $\lambda_{i,1}\leqslant\cdots\leqslant\lambda_{i,n-1}$ 是 $(n-1)\times(n-1)$ 主子矩阵 $A(\{i\})$ 的特征值，证明

$$\lambda_1\leqslant\lambda_{i,1}\leqslant\lambda_2\leqslant\lambda_{i,2}\leqslant\cdots\leqslant\lambda_{i,n-1}\leqslant\lambda_n.$$

这些交错不等式常归功于 Cauchy. 同时证明这些不等式蕴涵(4.3.15)中的诸不等式. **提示**：利用(4.3.8)或(4.3.15).

17. 如果 $A=[a_{ij}]\in M_n$ 是 Hermite 矩阵，且对某个 i，$a_{ii}=\lambda_n$. 证明，对所有 $k=1, 2, \cdots, n$，$k\neq i$，有 $a_{ik}=a_{ki}=0$；如果 $a_{ii}=\lambda_1$，则也有类似的结论. **提示**：对 $n>2$ 采用直接计算，然后应用交错性便可证明结论.

200

18. 设 $A\in M_n$ 是 Hermite 矩阵且设 $a_i=\det A(\{1,2,\cdots,i\})$，$i=1,2,\cdots,n$. 若 A 是非奇异的，证明 A 的非负特征值的个数等于序列 $+1, a_1, a_2, \cdots, a_n$ 中正负号的变化次数. 特别地，若这些主子式都是正的，则 A 就不会有负特征值. 如果某个 $a_i=0$，会出现什么情形? **提示**: 利用交错性.

19. 设 $A=[a_{ij}]\in M_n$ 是正规矩阵. 证明，若 A 有"小"列或行，则 A 一定有"小"特征值. 说得更明确些，设 A 的诸特征值的绝对值的平方 $\{|\lambda_i|^2: i=1,\cdots,n\}$ 按非减顺序排列，并且用 $v_1^2\leqslant v_2^2\leqslant v_3^2\leqslant\cdots\leqslant v_n^2$ 表示所得到的有序值. 设各行(或各列)元素的绝对值的平方和的平方根 $\{(\sum_{k=1}^{n}|a_{ik}|^2)^{1/2}: i=1,\cdots,n\}$ 按非减顺序排列，并且用 $R_1\leqslant R_2\leqslant\cdots\leqslant R_n$ 表示所得到的有序值. 证明

$$\sum_{i=1}^{k}v_i^2\leqslant\sum_{i=1}^{k}R_i^2,\quad 其中\ k=1,\cdots,n,$$

一个类似的上界可用列平方和表示. **提示**: 数量 v_i^2 是 Hermite 矩阵 AA^* 的特征值. AA^* 的诸主对角元是什么? 用优化和定理(4.3.26). 关于列和不等式可考虑 A^*A.

进一步阅读 关于优化的另外信息可参看[MOl]. 在定理(4.3.10)下面提到了诸主子矩阵的特征值间的一般交错不等式组，有关的讨论可参看 C. R. Johnson and H. A. Robinson, "Eigenvalue Inequalities for Principal Submatrices," *Lin. Alg. Appl.* 37(1981), 11-22. 习题4中概述的论断的原始证明可参看 H. Weyl, "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung," *Math. Annalen* 71 (1912), 441ff.；在文献的444-445(页)可看到其引理的证明. Weyl 用积分方程叙述并证明了他的结果，但把它转化成线性代数的形式是直接的.

4.4 复对称矩阵

矩阵 $A\in M_n$ 是对称的，是指 $A=A^T$. 在许多场合，所研究的对称矩阵只有实元素，因而它们是实 Hermite 矩阵，并且本章迄今所讨论的全部结果都适用于这些矩阵.

[201] 但是，在有些情形，我们要与复对称矩阵打交道. 一个例子是研究复平面中单位圆盘的正则解析映射，如果 $f(z)$ 是单位圆盘上的正则解析函数，又如果 $f(z)$ 是适合 $f(0)=0$ 和 $f'(0)=1$ 的标准化了的函数，那么，$f(z)$ 是一一的(有时称为单叶的)，当且仅当

$$\sum_{i,j=1}^{n}x_i\bar{x}_j\log\frac{1}{1-z_i\bar{z}_j}\geqslant\left|\sum_{i,j=1}^{n}x_ix_j\log\left[\frac{z_iz_j}{f(z_i)f(z_j)}\frac{f(z_i)-f(z_j)}{z_i-z_j}\right]\right|\tag{4.4.1}$$

对满足 $|z_i|<1$ 的点 $z_1,\cdots,z_n\in\mathbf{C}$ 的所有选择，点 $x_1,\cdots,x_n\in\mathbf{C}$ 的所有选择和所有 $n=1,2,\cdots$ 成立. 如果 $z_i=z_j$，则右边的差商可以看作 $f'(z_i)$. 这些称为 Grunsky 不等式组的庞杂不等式有很简单的代数形式

$$x^*Ax\geqslant|x^TBx|,\tag{4.4.2}$$

其中 $x=[x_i]\in\mathbf{C}^n$，$A=[a_{ij}]\in M_n$，$B=[b_{ij}]\in M_n$，

$$a_{ij}=\log\frac{1}{1-z_i\bar{z}_j},\quad b_{ij}=\log\left[\frac{z_iz_j}{f(z_i)f(z_j)}\frac{f(z_i)-f(z_j)}{z_i-z_j}\right].$$

应注意的是，A 是 Hermite 矩阵，而 B 是复对称矩阵.

另一个自然要产生复对称矩阵的例子出现在一般的矩问题中. 设$\{a_0, a_1, a_2, \cdots\}$是给定的复数序列, 设 $n\geqslant 1$ 是某个正整数, 且定义 $A_{2n}=[a_{ij}]\equiv[a_{i+j}]\in M_{2n}$, 注意 A_{2n}是形状为 Hankel 矩阵的复对称矩阵. 对 $x\in \mathbf{C}^{2n}$, 我们考虑复二次型 $x^TA_{2n}x$, 要问是否存在某个固定常数 $c>0$, 使得对所有 $x\in \mathbf{C}^{2n}$和所有 $n=1, 2, \cdots$有

$$|x^TA_{2n}x|\leqslant cx^*x.$$

根据 Nebari 定理, 这个条件成立, 当且仅当存在一个几乎处处有界的 Lebesgue 可测函数 $F(t)$: $F(t)$: $\mathbf{R}\to\mathbf{C}$, 它的 Fourier 系数是已知数 $a_0, a_1, a_2, \cdots$; 关于 $F(t)$的本质边界恰好是上述不等式组的常数 c.

在实际应用中复对称矩阵似乎不像复 Hermite(或实对称)矩阵那样几乎经常出现, 但是前两个例子说明, 它们还是出现了. 虽然复对称矩阵不一定可对角化(见本节末习题 15), 可是复对称矩阵有一个类似于 Hermite 矩阵的谱定理(4.1.5)的分解, 并且可以用逻辑上类似的方法来证明它. 我们首先证明一个与 Schur 三角分解定理(2.3.1)类似的定理, 它说明, 包括对称矩阵在内的一类矩阵总可以分解成 $A=U\Delta U^T$, 其中 U 是酉矩阵, Δ 是上三角矩阵, 如果上三角矩阵是对称的, 则它必定是对角矩阵. [202]

4.4.3 定理 设 $A\in M_n$ 是给定的, 那么存在酉矩阵 $U\in M_n$ 和上三角矩阵 $\Delta\in M_n$ 使得 $A=U\Delta U^T$, 当且仅当 $A\bar{A}$ 的所有特征值是非负实数, 在这个条件下, Δ 的所有主对角元可以选取非负值.

证明: 因为 U 是酉矩阵且 $U^T=\bar{U}^*$, 所以, 如果 $A=U\Delta U^T$, 则 $A\bar{A}=U\Delta U^T\bar{U}\bar{\Delta}U^*$. 当 Δ 是上三角矩阵时, 上三角矩阵 $\Delta\bar{\Delta}$ 的诸主对角元是非负实数, 且 $A\bar{A}$ 酉相似于$\Delta\bar{\Delta}$, 因此, 从上三角矩阵的诸特征值恰好是它的诸主对角元这一事实便可推出条件的必要性成立.

关于充分性, 假定 $A\bar{A}$ 只有非负特征值, 且设 x 是 $A\bar{A}$ 的一个特征向量; 即 $A\bar{A}x=\lambda x$, 且 $\lambda\geqslant 0$, $x\neq 0$. 有两种可能情形:

(a) $A\bar{x}$ 与 x 相关;

(b) $A\bar{x}$ 与 x 无关.

在前一种情形(a)(当 λ 是 $A\bar{A}$ 的单特征值时, 这种情形总是成立的), 存在某个 $\mu\in\mathbf{C}$, 使得 $A\bar{x}=\mu x$. 但是 $A\bar{A}x=A\overline{\mu x}=\bar{\mu}A\bar{x}=\bar{\mu}\mu x=|\mu|^2x=\lambda x$, 因而 $|\mu|^2=\lambda$. 在后一种情形(b)(如果 λ 是 $A\bar{A}$ 的重特征值, 这种情形可能成立), 对所有 $\mu\in\mathbf{C}$, 向量 $y=A\bar{x}+\mu x$ 是非零的, 并且可以选择 μ 为适合 $|\mu|^2=\mu\bar{\mu}=\lambda$ 的任一复数. 于是 $A\bar{y}=A(\bar{A}x+\overline{\mu x})=A\bar{A}x+\bar{\mu}A\bar{x}=\lambda x+\bar{\mu}A\bar{x}=\mu\bar{\mu}x+\bar{\mu}A\bar{x}=\bar{\mu}(A\bar{x}+\mu x)=\bar{\mu}y$. 在情形(a)或(b), 我们已证明, 存在某个非零向量 $v\in\mathbf{C}$ 和某个有 $|a^2|=\lambda$ 的 $a\in\mathbf{C}$, 使得 $A\bar{v}=av$, 因为这个恒等式在 v 乘以正纯量后不变, 所以还可以假定 v 是单位向量. 同时, 对任意 $\theta\in\mathbf{R}$, 有 $e^{-i\theta}A\bar{v}=A(\overline{e^{i\theta}v})=e^{-i\theta}av=(e^{-2i\theta}a)(e^{i\theta}v)$, 且当 v 是单位向量时, $e^{i\theta}v$ 也是单位向量, 因为可以选取 θ 使得 $e^{-2i\theta}a\geqslant 0$, 因而得出, 如果 $A\in M_n$, 且 λ 是 $A\bar{A}$ 的非负特征值, 则存在单位向量 v, 使得 $A\bar{v}=\sigma v$, 且 $\sigma=+\sqrt{\lambda}\geqslant 0$.

现在把这个向量 v 扩充为 $\mathbf{C}^n$ 的标准正交基$\{v_1, v_2, \cdots, v_n\}$, 且设 V_1 是以这些向量为列的矩阵. 因为标准正交性和等式 $A\bar{v}=\sigma v$, 矩阵 $V_1^*A\bar{V}_1$ 的第一列有元素 $v_i^*A\bar{v}=\sigma v_i^*v=\sigma\delta_{i1}$. 因此, 除了 $\bar{V}_1^TA\bar{V}_1$ 的第一列中第一个元素以外, 所有元素必须为零(第一个元素也可能是零).

[203] 如果用分块形式把这个矩阵写成

$$\bar{V}_1^T A V_1 = \begin{bmatrix} \sigma & w^T \\ 0 & A_2 \end{bmatrix}, \quad w \in \mathbf{C}^{n-1}, \quad A_2 \in M_{n-1}, \quad \sigma \geq 0, \tag{4.4.3a}$$

我们看到

$$(V_1^T A V_1)(\overline{V_1^T A V_1}) = V_1^* A\overline{A} V_1 = \begin{bmatrix} \sigma^2 & \sigma\bar{w}^T + w^T\overline{A}_2 \\ 0 & A_2\overline{A}_2 \end{bmatrix}.$$

因此，$A\overline{A}$ 的诸特征值(根据假定它们都是非负的)是 σ^2 以及 $A_2\overline{A}_2$ 的诸特征值. 由此得出，通过这个简化过程所得到的矩阵 $A_2 \in M_{n-1}$ 也有使 $A_2\overline{A}_2$ 的所有特征值都非负的性质.

现在可以对于 A_2 及其后继矩阵重复实施上述简化过程，至多经 $n-1$ 次(正像在 Schur 三角化定理(2.3.1)的证明中所做的那样)便得到

$$\bar{V}_{n-1}^T \cdots V_2^T V_1^T A V_1 V_2 \cdots V_{n-1} = \begin{bmatrix} \sigma_1 & & * \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix} = \Delta,$$

其中，Δ 是具有非负主对角元 σ_i 的上三角矩阵. 如果令 $U = V_1 V_2 \cdots V_{n-1}$，则有 $A = U\Delta U^T$，这正是所欲求的. □

练习 直接对矩阵 $A = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$ 实施定理(4.4.3)证明中的计算，并证明 $A = U\Delta U^T$，其中

$$\Delta = \begin{bmatrix} 0 & 2i \\ 0 & 0 \end{bmatrix}, \quad U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}.$$

如果 $n \geq 2$，并非每个矩阵 $A \in M_n$ 都有使 $A\overline{A}$ 的所有特征值均非负的性质：$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ 就是一个简单的例子. 因而，定理(4.4.3)与 Schur 三角化定理(2.3.1)只是部分类似. 每个 $A \in M_n$ 可经形如 $A \to UAU^*$ 的变换(其中酉矩阵 $U \in M_n$)三角化，不过，只有使 $A\overline{A}$ 有全部非负特征值的那些矩阵 $A \in M_n$ 可经形如 $A \to UAU^T$ 的变换(其中酉矩阵 $U \in M_n$)三角化.

每个对称矩阵 $A \in M_n$ 有如下性质：$A\overline{A} = AA^*$ 的所有特征值都是非负的. 该特殊形式已包含在定理(4.4.3)中，人们通常把它归功于 Schur (1945). 但是较早的证明是由 Hua (1944)，Siegel (1943) 和 Jacobsen (1939) 提出的；而历史的优先权显然应该属于 Takagi (1925).

[204] **4.4.4 推论**(Takagi 分解) 如果 $A \in M_n$ 是对称矩阵($A = A^T$)，则存在酉矩阵 $U \in M_n$ 和非负实对角矩阵 $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$ 使得 $A = U\Sigma U^T$. U 的诸列是由 $A\overline{A}$ 的特征向量组成的标准正交组，而 Σ 的相应对角元是 $A\overline{A}$ 的相应特征值的非负平方根.

证明：如果 $A = A^T$，则 $\overline{A} = A^*$，且 $A\overline{A} = AA^*$. 如果 $x \neq 0$ 是 Hermite 矩阵 AA^* 的任一特征向量，且 $AA^*x = \lambda x$，则 $x^*\lambda x = \lambda(x^*x) = x^*AA^*x = (A^*x)^*(A^*x)$. 因为 $y^*y \geq 0$ 对所有 $y \in \mathbf{C}^n$ 成立，而 $y^*y = 0$ 当且仅当 $y = 0$，所以有 $\lambda = (A^*x)^*(A^*x)/x^*x \geq 0$. 因此，只要 A 是对称矩阵，$A\overline{A}$ 的所有特征值都为负. 定理(4.4.3)保证存在酉矩阵 $U \in M_n$ 和上三角矩阵

$\Delta \in M_n$，其中

$$\Delta = \begin{bmatrix} \sigma_1 & & * \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}, \quad \text{所有 } \sigma_i \geqslant 0,$$

使得 $A = U\Delta U^T$. 但是 $U\Delta U^T = A = A^T = U\Delta^T U^T$，因而 $\Delta = \Delta^T$，这只有在 $\Delta \equiv \Sigma$ 是对角矩阵时才能成立，根据构造，Δ 是非负的. 最后，$A\bar{A} = U\Sigma U^T \bar{U}\Sigma U^* = U\Sigma^2 U^*$ 是 Hermite 矩阵 $A\bar{A}$ 的酉对角化，因而，U 的诸列是 $A\bar{A}$ 的特征向量. □

形如 $U\Lambda U^T$ 的任一矩阵，其中 Λ 是对角矩阵(不一定非负)，显然是对称的，因此，为了使某个矩阵 $A \in M_n$ 能分解成 $A = U\Lambda U^T = U\Lambda\bar{U}^* = U\Lambda\bar{U}^{-1}$，其中，$U$ 是酉矩阵，而 Λ 是对角矩阵，其必要充分条件是 A 是对称矩阵. 在定理(4.6.11)中给出了 A 可以分解成 $A = S\Lambda S^{-1}$ 的条件，其中 Λ 是对角矩阵，而 S 是非奇异矩阵(但不一定是酉矩阵).

每个复矩阵 $A \in M_n$ 可以写成形式 $A = V\Sigma W^*$，其中，V，$W \in M_n$ 是酉矩阵，而 Σ 是具有非负主对角元的对角矩阵. 这就是奇异值分解，将在(7.3)节中讨论它. Σ 的诸对角元是 A 的奇异值. 关于(可能是复)对称矩阵的 Takagi 分解 $A = U\Sigma U^T$ 是关于对称矩阵的特殊奇异值分解，其中 $V = \bar{W}$.

在定理(4.4.3)的证明中所采用的构造性方法可以用来计算复对称矩阵的 Takagi 分解. 因为 A 的对称性，所产生的矩阵 Δ 将自然是对角矩阵. 见本节末习题 9.

练习 直接对矩阵 $A = \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ 实施定理(4.4.3)证明中的计算，并证明 $A = U\Delta U^T$，其中，

205

$$\Delta = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}, \quad U = \frac{1}{\sqrt{4+2\sqrt{2}}} \begin{bmatrix} 1+\sqrt{2} & i \\ i & 1+\sqrt{2} \end{bmatrix}.$$

注意，Δ 自然是对角矩阵.

由于 Takagi 分解 $A = U\Sigma U^T$ 中酉因子 U 的诸列是 Hermite 矩阵 $A\bar{A}$ 的特征向量，这可能会误认为，如果 $A\bar{A} = U\Sigma^2 U^*$ 是酉对角化的，则 $A = U\Sigma U^T$. 实际情况不一定如此. 考察例子 $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 便可看出这一点. 因为 $A\bar{A} = I$，所以对任何 2×2 实正交矩阵 Q，有 $A\bar{A} = QI^2Q^T$，而 $QIQ^T = I \neq A$. 问题是 $A\bar{A}$ 有重数大于 1 的特征值，因而 $A\bar{A}$ 的任一特征向量都不可能有 $A\bar{x} = ax$ 的性质；这个特征向量不可能给出 A 的所欲求的化简. 如果考虑基向量 e_1，则 $A\bar{A}e_1 = Ie_1 = Ie_1$，而 $A\bar{e}_1 = Ae_1 = e_2$；于是有定理(4.4.3)证明中的情形(b). 根据证明，可以取 $w = A\bar{e}_1 + 1e_1 = e_2 + e_1$，这便得到向量 $v = v_1 = (e_1 + e_2)/\sqrt{2}$，它能化简 A. 因为 $v_2 = (e_1 - e_2)/\sqrt{2}$ 与 v_1 正交，可以取

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

并且得到 $V^TAV = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}^2 = \Sigma D^2$. 因此，如果令

$$U = VD = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix},$$

则 $A = UIU^T$ 是 A 的一个适当分解. 应指出的是，实对称矩阵的 Takagi 分解(4.4.4)不可能有实因子.

刚才所讨论的例子中的困难以及在一般情形中的困难是由 $A\bar{A}$ 的重特征值引起的. 如果 $A\bar{A}$ 的所有特征值是不同的，又如果采用(4.4.3)证明中的构造法来计算复对称矩阵 A 的 Takagi 分解，那么，总有情形(a)(见习题 9). 在这种情形，$A\bar{A}$ 的每个特征向量 x 有性质：对满足 $a = \sigma e^{2i\theta}$，$\theta \in \mathbf{R}$ 和 $A\bar{A}x = \sigma^2 x$ 的某个 $a \in \mathbf{C}$，有 $A\bar{x} = ax$. 因而，如果 $A\bar{A} = V\Sigma^2 V^*$ 是 Hermite 矩阵 $A\bar{A}$ 的酉对角化，则必有 $A\bar{V} = V\Sigma D^2$，其中 $D^2 = \mathrm{diag}(e^{2i\theta_1}, \cdots, e^{2i\theta_n})$；只要知道了 V 和 Σ(Σ^2 的非负平方根)，这个恒等式就可用来计算相应于 Σ 的非零对角元的 D^2 的对角元. 相应于 Σ 的零对角元的 D^2 的对角元是任意的，且可以取为 $+1$. 最后，如果令 $U = VD$ 和 $D = \mathrm{diag}(e^{i\theta_1}, \cdots, e^{i\theta_n})$，则有 $A = A\bar{V}V^T = V\Sigma D^2 V^T = (VD)\Sigma(VD)^T = U\Sigma U^T$. 我们把这些论断总结成下述推论.

206

4.4.5 推论 如果 $A \in M_n$ 是对称矩阵，且 $A\bar{A}$ 的特征值互不相同，又如果 $A\bar{A} = V\Sigma^2 V^*$ 是 $A\bar{A}$ 的酉对角化，其中 $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$ 且所有 $\sigma_i \geq 0$，则存在对角矩阵 $D = \mathrm{diag}(e^{i\theta_1}, \cdots, e^{i\theta_n})$(其中所有 $\theta_i \in \mathbf{R}$)使得 $A = U\Sigma U^T$(其中 $U = VD$). 相应于 Σ 的非零对角元的因子 D 的对角元由关系式 $A\bar{V} = V\Sigma D^2$ 确定；相应于 Σ 的零对角元的 D 的对角元可以取为 $+1$.

如果 $A \in M_n$ 是对称矩阵，并且利用(4.4.4)把 A 写成 $A = U\Sigma U^T$，也可以把它写成 $A = (U\Sigma^{1/2})(U\Sigma^{1/2})^T$，其中 $\Sigma^{1/2} = \mathrm{diag}(+\sqrt{\sigma_1}, +\sqrt{\sigma_2}, \cdots, +\sqrt{\sigma_n})$. 这番论证构成下述推论的证明.

4.4.6 推论 设 $A \in M_n$，则 A 是对称矩阵，当且仅当存在矩阵 $S \in M_n$ 使得 $A = SS^T$. 可以选取 $S = UD$，其中 U 是酉矩阵，$D = \mathrm{diag}(\sqrt{\sigma_1}, \sqrt{\sigma_2}, \cdots, \sqrt{\sigma_n})$，而 σ_i 是 A 的奇异值，在这种情形，rank S = rank A.

虽然实对称矩阵是正规的，但非实复对称矩阵不一定是正规的. 如果 $A = B + iC \in M_n$，其中 B 和 C 是实矩阵，则 A 是对称矩阵，当且仅当 B 和 C 都是实对称矩阵. 如果 A 既是对称矩阵，又是正规矩阵，则

$$AA^* = (B^2 + C^2) + i(CB - BC) = (B^2 + C^2) + i(BC - CB) = A^*A$$

由此可以推出 B 与 C 可交换. 在这种情形下，B 和 C 可经实正交矩阵 Q 同时对角化. 如果 $B = QD_1Q^T$，且 $C = QD_2Q^T$，其中 D_1 和 D_2 是实对角矩阵，则 $A = B + iC = QD_1Q^T + iQD_2Q^T = Q(D_1 + iD_2)Q^T = Q\Lambda Q^T$，其中 $\Lambda = D_1 + iD_2$. 反之，如果矩阵 $A \in M_n$ 可写成 $A = Q\Lambda Q^T$，其中 Q 是实正交矩阵，而 Λ 是对角矩阵，则 $A = A^T$，且 $AA^* = Q\Lambda Q^T Q\bar{\Lambda} Q^T = Q|\Lambda|^2 Q^T = Q\bar{\Lambda}Q^T Q\Lambda Q^T = A^*A$，因而 A 既是对称矩阵，又是正规矩阵. 这就证明了下面的定理.

4.4.7 定理 设 $A \in M_n$，那么，A 既是对称矩阵，又是正规矩阵，当且仅当存在实正交矩阵 $Q \in M_n(\mathbf{R})$ 和对角矩阵 $\Lambda \in M_n$ 使得 $A = Q\Lambda Q^T$.

一个既对称又正规的复矩阵的简单的有用例子是

$$S = \frac{1}{\sqrt{2}}(I + iB), \tag{4.4.8}$$

其中 B 是“后向单位”矩阵 [207]

$$B = \begin{bmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{bmatrix},$$

它曾在(3.2.3)中证明每个矩阵相似于它的转置时起过作用.

因为 $B^2 = I$，所以

$$S\overline{S} = \frac{1}{2}(I + iB)(I - iB) = \frac{1}{2}(I - iB + iB + B^2) = I,$$

由此可知，S 既是对称矩阵又是酉矩阵.

现在考虑具有零主对角线的标准 Jordan 块 $J_k(0)$，$k \geqslant 2$，把它写成形式

$$N = \begin{bmatrix} 0 & 1 & & 0 \\ & & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix} \in M_k.$$

经简单计算可知

$$BNB = \begin{bmatrix} 0 & & & 0 \\ 1 & & & \\ & & \ddots & \\ & \ddots & & \\ 0 & & 1 & 0 \end{bmatrix},$$

$$BN = \begin{bmatrix} 0 & & & 0 \\ & & & 1 \\ & & \iddots \; \iddots & \\ 0 & 1 & & 0 \end{bmatrix},$$

$$NB = \begin{bmatrix} 0 & & 1 & 0 \\ & \iddots & & \\ 1 & \iddots & & \\ 0 & & & 0 \end{bmatrix}.$$

因此，N 酉相似于矩阵

$$\begin{aligned} SNS^{-1} = SN\overline{S} &= \frac{1}{2}(I + iB)N(I - iB) \\ &= \frac{1}{2}(N + BNB) + \frac{i}{2}(BN - NB) \\ &= \frac{1}{2}\begin{bmatrix} 0 & 1 & & 0 \\ 1 & & \ddots & \\ & \ddots & & \\ & \ddots & & 1 \\ 0 & & 1 & 0 \end{bmatrix} + \frac{i}{2}\begin{bmatrix} 0 & & -1 & 0 \\ & \iddots & & 1 \\ 1 & & \iddots \; \iddots & \\ 0 & 1 & & 0 \end{bmatrix}, \end{aligned} \tag{4.4.8a}$$

[208]

它显然是对称矩阵．任一 Jordan 块 $J_k(\lambda)(k\geqslant 2)$ 具有形式 $\lambda I+N$，又 SNS^{-1} 是对称矩阵，所以 $SJ_k(\lambda)S^{-1}=S(\lambda I+N)S^{-1}=\lambda I+SNS^{-1}$ 是对称矩阵．

每个矩阵 $A\in M_n$ 都相似于它的 Jordan 标准形 J(3.1.14)，其中 $\epsilon=2$，且 $J=J_{n_1}(\lambda_1,2)\oplus\cdots\oplus J_{n_k}(\lambda_k,2)$ 是修改后的诸 Jordan 块 $J_{n_i}(\lambda_i,2)$ 的直和．这个论断(它相当于在上述论证中用 $2N$ 代替 N)允许我们略去(4.4.8a)中的系数因子 $\frac{1}{2}$．当 $n_i\geqslant 2$ 时，如果我们设 $S_{n_i}=(1/\sqrt{2})(I+iB)\in M_{n_i}$ 是形如(1.4.8)的 $n_i\times n_i$ 矩阵，且 $S_1=[1]$；如果令 $T=S_{n_1}\oplus\cdots\oplus S_{n_k}$，则上述论证说明，

$$TJT^{-1}=TJ\overline{T}=(S_{n_1}J_{n_1}(\lambda_1,2)\overline{S}_{n_1})\oplus\cdots\oplus(S_{n_k}J_{n_k}(\lambda_k,2)\overline{S}_{n_k})$$

是对称矩阵的直和，因而是对称矩阵．因为每个 S_{n_i} 是酉矩阵，所以矩阵 T 是酉矩阵，因此已经证明，呈 Jordan 标准形的每个矩阵等价于一个对称矩阵．因为每个矩阵相似 Jordan 矩阵，故已经证明了下述定理．

4.4.9 定理 每个矩阵 $A\in M_n$ 相似于对称矩阵．

实际上，已经证明每个矩阵 $A\in M_n$ 相似于对称 Jordan 标准形 $S_{n_1}(\lambda_1)\oplus\cdots\oplus S_{n_k}(\lambda_k)$，其中，如果 $\lambda=\alpha+i\beta$，且 $\alpha,\beta\in\mathbf{R}$ 则

$$S_k(\lambda)=SJ_k(\lambda,2)\overline{S}=\lambda I+SN\overline{S}$$

$$=\lambda I+\begin{bmatrix}0&1&&0\\1&0&\ddots&\\&\ddots&\ddots&1\\0&&1&0\end{bmatrix}+i\begin{bmatrix}&0&-1&0\\&\iddots&\iddots&1\\-1&&\iddots&\\0&1&&0\end{bmatrix}\in M_k,$$

且 S 由(4.4.8)给出．注意

$$S_1(\lambda)=[\lambda]\quad\text{而}\quad S_2(\lambda)=\begin{bmatrix}\lambda-i&1\\1&\lambda+i\end{bmatrix}.$$

209 因为这种形式是直接从 Jordan 标准形推导出来的，所以它的唯一性与 Jordan 标准形的唯一性相同．

这个结论的一个推论是，关于复对称矩阵的谱、Jordan 块、极小多项式、特征多项式或不变因式都没有任何特别的结果．如果这些量中任何一个可以出现在某个阶数的对称矩阵中，则它也可以出现在同阶的一般复矩阵中．M_n 中的每个相似类包含一个对称矩阵，$\mathbf{C}^n$ 上的每个线性变换有一个对称的基表示，矩阵的对称性只不过是为表示相应线性变换而选定一个特殊的基的人为现象．另一个推论是，每个矩阵在某种意义下"可对角化"．

4.4.10 推论 设 $A\in M_n$ 已知．则存在非奇异矩阵 S 和酉矩阵 U，使得 $(US)A(\overline{U}S)^{-1}$ 是具有非负对角元的对角矩阵．

证明： 利用(4.4.9)求非奇异矩阵 $S\in M_n$ 使 SAS^{-1} 是对称矩阵，然后利用(4.4.4)求酉矩阵 $U\in M_n$ 使 $U(SAS^{-1})U^T$ 是非负对角矩阵． □

定理(4.4.9)同时推出：每个复矩阵相似于它的转置，且每个复矩阵可以写成两个复对称矩阵的乘积．这两个结论对任意域上的矩阵都成立，但定理(4.4.9)对一般域不成立．

4.4.11 推论 设 $A\in M_n$ 已知. 则存在矩阵 B, $C\in M_n$, 使得 $B=B^T$, $C=C^T$, 且 $A=BC$. B 或 C 可以选为非奇异矩阵.

证明: 利用本定理把 A 写成 $A=SES^{-1}$, 其中 $E=E^T$, 且 S 是非奇异矩阵. 于是 $A=(SES^T)(S^T)^{-1}S^{-1}=(SES^T)(SS^T)^{-1}=BC$, 其中 $B=SES^T$ 和 $C=SS^T$ 都是对称矩阵. 又因为 $A=(SS^T)(S^{-1})^TES^{-1}$, 所以因子 B 或 C 可以选为非奇异矩阵. □

Gram-Schmidt 过程(0.6.4)在研究正规矩阵中有许多应用. 有一个类似的过程对研究复对称矩阵是有用的.

4.4.12 引理 设 x_1, …, $x_k\in\mathbf{C}^n$ 是给定的向量, 且 $k\leqslant n$. 则存在向量 y_1, …, y_k, 使得 $\text{Span}\{x_1, \cdots, x_k\}=\text{Span}\{y_1, \cdots, y_k\}$, 并且对所有 i, $j=1$, 2, …, k 及 $i\neq j$ 有 $y_i^Ty_j=0$, 对 $i=1$, 2, …r 有 $y_i^Ty_i=1$, 对 $i=r+1$, …, k 有 $y_i^Ty_i=0$. 其中 $r=\text{rank } X^TX$ 且 $X=[x_1\cdots x_k]\in M_{n,k}$ 是其列为已知向量$\{x_i\}$的矩阵. [210]

证明: 因为矩阵 X^TX 是对称矩阵, Takagi 分解定理(4.4.4)使我们可以把它写成 $X^TX=U\Sigma U^T$, 其中, $U\in M_k$ 是酉矩阵且 $\Sigma=\text{diag}(\sigma_1, \cdots, \sigma_k)$, 而 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>\sigma_{r+1}=0=\cdots=\sigma_k$, $\text{rank } X^TX=r$. 如果令 $D=\text{diag}(\sqrt{\sigma_1}, \cdots, \sqrt{\sigma_r}, 1, \cdots, 1)\in M_k$, 且记 $I_r=\text{diag}(1, \cdots, 1, 0, \cdots, 0)\in M_k$, 它有 r 个 1 和 $k-r$ 个 0, 则 $X^TX=(UD)I_r(UD)^T=S^TI_rS$ 其中 $S=DU^T$ 是非奇异矩阵. 因而$(XS^{-1})^T(XS^{-1})=I_r$. 因此, 如果令 $XS^{-1}\equiv Y=[y_1, \cdots y_k]\in M_{n,k}$, 因为 $Y^TY=I_r$. 则列向量 y_1, …, y_n 具有所要证明的性质. □

上述引理叙述的法则形式上类似于 Gram-Schmidt 过程, Gram-Schmidt 过程是针对 X^*X 而不是 X^TX. 在 Gram-Schmidt 过程中, 对每个 $j=1$, 2, …, k, 每个 y_j 可以作成 x_1, …, x_j 的线性组合, 但在这里可能行不通. 另一个差别是, 在 Gram-Schmidt 过程中, 具有 $y_i^*y_i=1$ 的向量 y_i 的个数等于 rank X(诸无关向量 x_i 的最大个数), 它等于 rank X^*X. 但是在这种情形, 具有 $y_i^Ty_i=1$ 的向量 y_i 的个数等于 X^TX 的秩, 它可能小于 rank X.

例 假定 $k=1$, 且 $x_1=X=\begin{bmatrix}1\\ i\end{bmatrix}$. 则 $X^TX=0$, 因而 $0=\text{rank } X$, 它严格小于 $\text{rank } X=1$, y_1 只可能是 x_1 的纯量倍, 因此不可能选取 y_1 使得 $\text{Span}\{x_1\}=\text{Span}\{y_1\}$而又 $y_1^Ty_1=1$.

例 假定 $k=2$, 且 $X=[x_1x_2]=\begin{bmatrix}1 & i\\ i & 1\end{bmatrix}$. 则 $\text{rank } X^TX=2$, 并且存在向量 y_1, y_2 使得 $\text{Span}\{y_1, y_2\}=\text{Span}\{x_1, x_2\}$及 $y_1^Ty_1=1=y_2^Ty_2$. 因为 $x_1^Tx_1=0$, 所以不可能选取 y_1 为 x_1 的纯量倍.

我们所考虑的直接应用是针对可对角化复对称矩阵的特殊情形的. 如果 $A=A^T\in M_n$, 且 $A=S\Lambda S^{-1}$. 其中, 对角矩阵 $\Lambda\in M_n$, 非奇异矩阵 $S\in M_n$, 则显然不能从这个通常的对角化表示推出 A 是对称矩阵. 但是, 如果 S 是复正交矩阵, 则 $S^{-1}=S^T$, 且 $A=S\Lambda S^{-1}=S\Lambda S^T$ 显然是对称矩阵. 下面的定理说明, 总可以选取 S 为复正交矩阵.

4.4.13 定理 设 $A\in M_n$ 是对称矩阵, 则 A 可对角化, 当且仅当它可复正交对角化, 也就是说, $A=S\Lambda S^{-1}$对于对角矩阵 $\Lambda\in M_n$ 和非奇异矩阵 $S\in M_n$ 成立当且仅当 $A=Q\Lambda Q^T$, 其中 $Q\in$ [211]

M_n 适合 $Q^TQ=I$.

证明：假定 $A=A^T$，又设 x，$y\in\mathbf{C}^n$ 是 A 的特征向量，且 $Ax=\lambda x$，$Ay=\mu y$. 如果 $\lambda\neq\mu$，则 $y^TAx=y^T\lambda x=\lambda y^Tx$，且 $y^TAx=(Ay)^Tx=(\mu y)^Tx=\mu y^Tx$，因而 $\lambda y^Tx=\mu y^Tx$，又因为 $\lambda\neq\mu$，所以 $y^Tx=0$. 这只不过是把双正交性原理(1.4.7)应用于对称矩阵. 如果 A 是可对角化的，且 $A=S\Lambda S^{-1}$，不失一般性，假定在 $\Lambda=\Lambda_1\oplus\cdots\oplus\Lambda_d$ 中，Λ 的相同特征值排放在一起，其中 $\Lambda_i=\lambda_iI\in M_{n_i}$，$n_1+\cdots+n_d=n$，且当 $i\neq j$ 时 $\lambda_i\neq\lambda_j$. 把 S 的诸列块分成 $S=[s_1\cdots s_n]=[S_1S_2\cdots S_d]$ 使之与 $\Lambda=\Lambda_1\oplus\cdots\oplus\Lambda_d$ 分法相同，于是，对 $i=1$，2，…，d，$S_i\in M_{n,n_i}$. 由双正交性质，如果 $i\neq j$，则 $S_i^TS_j=0\in M_{n_i,n_j}$. 因为 S^TS 是非奇异分块对角矩阵，所以，对所有 $i=1$，2，…，d，$S_i^TS_i$ 非奇异. 由于每个矩阵 $S_i^TS_i$ 是满秩的，引理(4.1.12)说明，每个 S_i 的诸列可以用新列来代替，它们是诸旧列的非奇异线性组合，且彼此复正交；即存在非奇异矩阵 $R_i\in M_{n_i}$ 使 $Q_i=S_iR_i$ 适合 $Q_i^TQ_i=R_i^TS_i^TS_iR_i=I\in M_{n_i}$. 因为对所有 $i\neq j$，$Q_i^TQ_j=R_i^TS_i^TS_jR_j=0$，又对 $i=1$，2，…，d，$AQ_i=AS_iR_i=\lambda_iS_iR_i=\lambda_iQ_i$，所以矩阵 $Q=[Q_1\cdots Q_d]\in M_n$ 是复正交的，且 $A=Q\Lambda Q^T$. □

上述结论可对定理(4.4.7)作出很好的解释：对称矩阵 A 可对角化，当且仅当 $A=Q\Lambda Q^T$ 且 Q 是复正交矩阵；又 A 是正规矩阵，当且仅当 Q 可以选为实正交矩阵.

可以对定理(4.1.13)中的结果稍作推广. 若 A，$B\in M_n$ 是对称矩阵，则 A 与 B 相似当且仅当它们可以通过复正交相似来实现相似. 事实上，假定存在一个多项式 $p(t)$ 使得 $A^T=p(A)$ 且 $B^T=p(B)$，在这个较弱的假定下上述推广成立. 见[HJ].

习题

1. 假定 $A\in M_n$ 是对称矩阵，且 $A=B+iC$，其中 B，$C\in M_n$ 都是实矩阵. 证明，A 是正规矩阵，当且仅当 B 与 C 可交换. 证明，A 是正规矩阵，当且仅当 $A\bar{A}$ 是实矩阵. 证明，A 是正规矩阵，当且仅当 A 与 $\bar{A}$ 可交换. 给出一个对称矩阵不是正规矩阵的例子.

2. 下面给出推论(4.4.4)的另一个证明的要点，请作详细的论述. 记号和假设如(4.4.4)中所述. 若 A 是奇异的，设$\{u_1, \cdots, u_k\}$是 A 的零空间的一组标准正交基且 $U=[u_1\cdots u_ku_{k+1}\cdots u_n]\in M_n$ 是酉矩阵. 则

212

$$U^TAU=\begin{bmatrix}0 & 0\\ 0 & A'\end{bmatrix},\quad A'\in M_{n-k},$$

其中 A'是非奇异对称矩阵. 因此，不失一般性，我们可以假定 A 是非奇异的. 设 $A=B+iC$，其中 B，C 是实矩阵，且设 $z=x+iy\in\mathbf{C}^n$，其中 x，$y\in\mathbf{R}^n$. 又设 $F=\begin{bmatrix}B & C\\ C & -B\end{bmatrix}$，$\hat{z}=\begin{bmatrix}x\\ -y\end{bmatrix}\in\mathbf{R}^{2n}$. (a)$B$，$C$ 和 F 是实对称矩阵. 讨论 $Az=(B+iC)(x+iy)$与 $F\hat{z}$ 之间的关系. (b)F 是非奇异的. **提示**：若 $F\hat{z}=0$，Az 是什么？(c)若 $F\begin{bmatrix}x\\ -y\end{bmatrix}=\lambda\begin{bmatrix}x\\ -y\end{bmatrix}$，则 $F\begin{bmatrix}y\\ x\end{bmatrix}=-\lambda\begin{bmatrix}y\\ x\end{bmatrix}$. 可以把 F 的非零特征值按一正一负配对. (d)设 F 相应于其正特征值λ_1，…，λ_n 的标准正交特征向量记作 $\hat{z}_i=\begin{bmatrix}x_i\\ -y_i\end{bmatrix}\in\mathbf{R}^{2n}$，$i=1$，2，…，$n$，设 $X=[x_1\cdots x_n]$，$Y=[y_1\cdots y_n]\in M_n$，又设 Σ

$=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)\in M_n$. 关于实对称矩阵的谱定理是指 $F=V\Lambda V^T$, 其中

$$V=\begin{bmatrix} X & Y \\ Y & X \end{bmatrix} \quad 而 \quad \Lambda=\begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix},$$

且 V 是实正交矩阵.(为什么?)设 $U=X-iY$. 证明 U 是酉矩阵且 $U\Sigma U^T=A$.

3. 当 A 是实对称矩阵时,(4.4.4)说的是什么? 它与实对称矩阵的通常谱分解有何关系? **提示**: 如果 $A=Q\Lambda Q^T$, 其中, Λ 是实对称矩阵, 而 Q 是实正交矩阵, 把 Λ 写成 $\Lambda=\Sigma D^2$ 且设 $U=QD$. 什么时候 Takagi 分解 $A=U\Sigma U^T$ 中的所有因子可取实矩阵?

4. 如果 $A=U\Sigma U^T\in M_n$, 且 U 和 Σ 如(4.4.4)中所述, 试通过直接计算证明, σ_i^2 是 $\bar{A}A$ 和 $A\bar{A}$ 的特征值, 且 $\bar{A}A$ 和 $A\bar{A}$ 是 Hermite 矩阵. 证明 U 的列 u_i 和数 σ_i 适合方程 $A\bar{u}_i=\sigma_i u_i$, $i=1, 2, \cdots, n$. 或许因为这个理由, 有时称 σ_i 为广义特征值, 不过, 术语奇异值似乎更为通用.

5. 假定 $A\in M_n$ 是对称矩阵, 设 Σ 和 U 如(4.4.4)中所述, 且把 A 的诸奇异值 σ_i 排成递减顺序 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n\geqslant 0$. (a)试修改 Rayleigh-Ritz 定理(4.2.2)的证明来证明 $\sigma_{\max}=\sigma_1=\max\{|x^TAx|/x^*x: 0\neq x\in\mathbf{C}^n\}$, 即类似于(4.2.2)中的上界, 复对称矩阵有相应的论述. 试考虑 U 的第 1 列来证明, 该极值由适合 $A\bar{x}=\sigma_1 x$ 的单位向量来达到. (b)试考虑 $A=I\in M_2$ 和 $x=\begin{bmatrix}1\\ i\end{bmatrix}$ 来说明, 在这种情形, $\sigma_{\min}=\sigma_n\neq\min\{|x^TAx|/x^*x: 0\neq x\in\mathbf{C}^n\}$, 因而, 类似于(4.2.2)中的下界, 对复对称矩阵的相应论述不成立. (c)试考虑 $A=I\in M_2$, $w=\begin{bmatrix}1\\ i\end{bmatrix}$ 来说明 $\max\{|x^TAx|/x^*x: 0\neq x\in\mathbf{C}^n, x\perp w\}=0$. 由此得出, 类似于 Courant-Fischer 极小-极大公式(4.2.12), 对复对称矩阵及其奇异值的相应论述当 $k>1$ 时不成立. 但是, 可以看一看(7.3.10). (d)类似于极大-极小公式(4.2.13), 关于对称矩阵的相应论述是什么? (e)设 $\tilde{A}\equiv\begin{bmatrix}1 & i\\ i & 1\end{bmatrix}$($\tilde{\sigma}_1=\tilde{\sigma}_2=\sqrt{2}$), 而 $A=[1]$($\sigma_1=1$)是删去 $\tilde{A}$ 的最后一行和最后一列后形成的矩阵. 注意, 类似于(4.3.9), 交错不等式 $\tilde{\sigma}_1\geqslant\sigma_1\geqslant\tilde{\sigma}_2$ 是不成立的. (f)不过, 还是有关于加边对称矩阵的诸奇异值的不等式. 设 $\tilde{A}\in M_{n+1}$ 是对称矩阵且有奇异值 $\tilde{\sigma}_1\geqslant\cdots\geqslant\tilde{\sigma}_{n+1}$, 而 $A\in M_n$(有奇异值 $\sigma_1\geqslant\cdots\geqslant\sigma_n$)是删去 $\tilde{A}$ 的一行和相应的列形成的矩阵. 试用定理(7.3.9)证明, $\tilde{\sigma}_k\geqslant\sigma_k\geqslant\tilde{\sigma}_{k+2}$, $k=1, \cdots, n$($\tilde{\sigma}_{n+2}\equiv 0$). 对于(e)中的例子验证这些不等式, 然后把它们与关于加边 Hermite 矩阵的诸特征值的诸交错不等式(4.3.9)进行比较.

213

6. 如果 $A\in M_n$ 是对称矩阵, 又如果 $A=U\Sigma U^T$, 其中, U 是酉矩阵, $\Sigma=\mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$, 且所有 $\sigma_i\geqslant 0$, 证明 A 的秩等于非零项 σ_i 的个数. **提示**: 如果 B, $C\in M_n$ 是非奇异矩阵, 则 $\mathrm{rank}\ A=\mathrm{rank}\ BAC$.

7. 设 $A=B+iC\in M_n$, 其中 B, C 是实矩阵, 又设 $F=\begin{bmatrix}B & C\\ C & -B\end{bmatrix}\in M_{2n}$. (a)证明, $\bar{A}A=B^2+C^2+i(BC-CB)$, 且

$$F^2=\begin{bmatrix} B^2+C^2 & BC-CB \\ -(BC-CB) & B^2+C^2 \end{bmatrix}.$$

(b) 证明 $S\equiv(1/\sqrt{2})\begin{bmatrix} I & -iI \\ -iI & I \end{bmatrix}\in M_{2n}$ 是酉矩阵. (c)证明 $SF^2S^*=\begin{bmatrix}\bar{A}A & 0\\ 0 & A\bar{A}\end{bmatrix}$.

(d) 证明 F 的诸特征值的平方就是 $\bar{A}A$ 的诸特征值及其复共轭. (e)如果 A 是复对称矩阵，证明，F 是具有实特征值的实对称矩阵，F^2 只有非负特征值，且 F 的特征值平方的集合与 Hermite 矩阵 $\bar{A}A$ 的特征值集合相同.

8. 设 $A\in M_n$ 是复对称矩阵. 考虑二次型 $q_A(x, x)=x^TAx$ 和由 A 生成的双线型 $b_A(x, y)=x^TAy$. 试用推论(4.4.4)证明

$$\sup_{x^*x=1} |q_A(x,x)| = \sup_{\substack{x^*x=1\\ y^*y=1}} |b_A(x,y)| = \sigma_{\max}(A),$$

[214] 其中 $\sigma_{\max}(A)$是 $\bar{A}A$ 的最大特征值.

9. 试用(4.4.3)的证明中的记号证明下列命题：(i)如果 λ 是 $A\bar{A}$ 的单特征值，且 $x\neq 0$ 适合 $A\bar{A}x=\lambda x$，则 $A\bar{x}$ 与 x 相关. **提示**：设 $\sigma=+\sqrt{\lambda}$且令 $w=A\bar{x}-\sigma x$. 证明 $A\bar{w}=-\sigma w$ 和 $A\bar{A}w=\lambda w$，因而 w 是 x 的纯量倍数. (ii)如果 $A=A^T$，则 $\bar{V}_1^TAV_1=[\sigma]\oplus A_2$；即(4.4.3a)中行向量 w^T 是零. 试用它来证明，这个构造法自然产生矩阵 $V_{n-1}^T\cdots\bar{V}_1^TA\bar{V}_1\bar{V}_2\cdots\bar{V}_{n-1}=U^TA\bar{U}=\Delta$，而 Δ 是对角矩阵.

10. 设 $A\in M_n$，且假定有非奇异矩阵 $S\in M_n$ 使得 $A=S\Lambda S^{-1}$，其中 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 证明，$A\bar{A}$ 可对角化，且它只有非负特征值，并且 rank $A=$rank $A\bar{A}$. 这与(4.1.4)有什么关系？说明$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$或$\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}$都不能写成这种形式.

11. 如果 $S\in M_n$ 是某个矩阵，证明，一般有 rank $S^TS\leqslant$rank S，也可能有 rank $S^TS<$rank S. 如果 S 是实矩阵，会出现什么情形？**提示**：考察 $S=\begin{bmatrix}1 & 0\\ i & 0\end{bmatrix}$.

12. 如果 $A\in M_n$ 是复对称矩阵，又 x，$y\in \mathbf{C}^n$ 是 A 的相应于 A 的不同特征值的特征向量，证明 $x^Ty=0$. 这说明 x 与 y 正交吗？**提示**：考虑 $x^T(Ay)=(Ax)^Ty$.

13. 如果 $A\in M_n$ 是对称矩阵，且有 n 个不同的特征值，直接证明存在非奇异矩阵 $S\in M_n$ 和对角矩阵 D 使得 $A=SDS^T$. **提示**：A 可对角化，因而 $A=S\Lambda S^{-1}$且 $AS=S\Lambda$. 根据习题 12，$S^TS=D$ 是对角矩阵. 因此 $S^TAS=S^TS\Lambda=D\Lambda$ 且 $A=(S^{-1})^T(D\Lambda)S^{-1}$. 为了证明有复正交矩阵 Q 使 $A=Q\Lambda Q^T$，需作哪些修改？

14. 如果 $A\in M_n$ 是非奇异对称矩阵，证明 A^{-1}是对称矩阵.

15. 实对称矩阵是 Hermite 矩阵，因而可对角化. 说明复对称矩阵不一定可对角化. **提示**：考察 $A=\begin{bmatrix}1 & i\\ i & -1\end{bmatrix}$并计算 A^2.

16. 设 $A\in M_n$，证明，A 是对称酉矩阵，当且仅当 A 可以写成 $A=Q\Lambda Q^T$，其中，$Q\in M_n(\mathbf{R})$是实正交矩阵，而 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)=\mathrm{diag}(e^{i\theta_1}, \cdots, e^{i\theta_n})$，并且对 $k=1, 2, \cdots, n$ 有 $|\lambda_k|=1$ 和 $\theta_k\in\mathbf{R}$.

17. 试用习题 16 证明，矩阵 $U\in M_n$ 是对称的酉矩阵，当且仅当存在酉矩阵 $V\in M_n$ 使得 $U=VV^T$.

[215]

18. 我们已经证明每个矩阵 $A\in M_n$ 相似于一个对称矩阵、每个矩阵相似于一个 Hermite

矩阵吗？相似于一个正规矩阵吗？

19. 利用(4.4.9)证明每个矩阵相似于它的转置.

20. 证明定理(4.4.9)在实数域上不成立；即不是每个矩阵 $A\in M_n(\mathbf{R})$都相似于实对称矩阵.

21. 复对称矩阵 A 可能有迷向向量 v 作为特征向量；即，$Av=\lambda v$，$v\neq 0$，且 $v^Tv=0$，但是，如果 A 可对角化，证明 λ 不可能是单特征值. 提示：一方面把 A 写成 $S\Lambda S^{-1}$，且 v 为 S 的第一列，另一方面证明，因为 S^TS 的第一行为零，所以 S^TS 是奇异矩阵. 特别是，如果 $v\in\mathbf{C}^n$ 是使 $v^Tv=0$ 的任一向量，则(秩 1)对称矩阵 $A=vv^T$ 不能对角化. 参看习题 15.

22. 对推论(4.4.4)的另一个证明的下述要点给出详细的论述. 记号和假设如(4.4.4)中所述. 这实质上是 Siegel (1943)的证明. (a)$A\bar{A}$ 是 Hermite 矩阵，因而存在一个酉矩阵 $V\in M_n$ 和一个实对角矩阵 $\Lambda_1\in M_n$ 使得 $A\bar{A}=V\Lambda_1V^*$. (b)$V^*A\bar{V}=B$ 既是对称矩阵又是正规矩阵，所以，根据(4.4.7)，存在一个对角矩阵 $\Lambda\in M_n$ 和一个实正交矩阵 $Q\in M_n(\mathbf{R})$使得 $B=Q\Lambda Q^T$. (c)$A=(VQ)\Lambda(VQ)^T$. 现在记 $\Lambda=E\Sigma E^T$，其中，E、Σ 均为对角矩阵，且 Σ 是非负矩阵，于是 $A=U\Sigma U^T$，其中 $U=VQE$ 是酉矩阵.

23. 设 $z=\{z_1, z_2, \cdots, z_n\}^T$ 是 n 个复变量的向量，又设 $f(z)$是在某个区域 $D\subset\mathbf{C}^n$ 中 n 个复变量的复解析函数. 因为混合偏导数相等，所以 $H=[\partial^2 f/\partial z_i\partial z_j]$在每一点 $z\in D$ 是对称矩阵. (4.0.3)中的讨论说明，可以假定一般线性偏微分算子

$$Lf=\sum_{i,j=1}^{n}a_{ij}(z)\frac{\partial^2 f}{\partial z_i\partial z_j}$$

中的系数矩阵 $A=[a_{ij}]$是对称的. 证明，在某一点 $z_0\in D$ 存在变量 $z\to U\zeta$ 的酉变换，使得在新坐标系下 Lf 在 z_0 是对角的，即

$$Lf=\sum_{i=1}^{n}\sigma_i\frac{\partial^2 f}{\partial\zeta_i^2},\quad 在\ z=z_0,\quad \sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n\geqslant 0.$$

24. 利用(4.4.13)以及类似于在(1.3.19)的证明中所采用的归纳证法，证明下述命题. 它与同时酉对角化一个 Hermite 矩阵族的定理(4.1.6)类似：设 $\mathcal{F}\subset M_n$ 是给定的可对角化对称矩阵族. 则对于所有 $A\in\mathcal{F}$，存在复正交矩阵 Q 使得 QAQ^T 是对角矩阵，当且仅当 $\mathcal{F}$ 是交换族. 216

25. 利用定理(4.4.7)证明中的证法证明，矩阵 $A\in M_n$ 既是斜对称的($A=-A^T$)又是正规的，当且仅当有实正交矩阵 $Q\in M_n(\mathbf{R})$使得 $Q^TAQ=0\oplus 0\oplus\cdots\oplus 0\oplus A_1\oplus A_2\oplus\cdots\oplus A_k$，其中每个 $A_j\in M_2$ 有形式

$$A_j=\begin{bmatrix}0 & z_j\\ -z_j & 0\end{bmatrix},\quad z_j\in\mathbf{C},\quad j=1,2,\cdots,k. \tag{4.4.14}$$

提示：考察 A 的实部和虚部，并利用定理(2.5.15). 什么时候 1×1 零直加项不出现？

26. 利用习题 25 以及习题 22 中的论断证明一个类似于复对称矩阵的 Takagi 分解(4.4.4)的复斜对称矩阵的分解：矩阵 $A\in M_n$ 是斜对称的($A=-A^T$)当且仅当存在一个酉矩阵 $U\in M_n$ 使得

$$A=U(0\oplus\cdots\oplus 0\oplus A_1\oplus\cdots\oplus A_k)U^T,$$

其中每个 $A_j\in M_n$ 有形式(4.4.14). 特别地，可得出一个斜对称复矩阵的秩一定是偶数.

27. 设 $W\in M_n$ 是给定的酉矩阵. 证明，只要 $A\in M_n$ 适合 $W^TA=AW$，就一定存在酉矩阵

$V\in M_n$ 使得 $V^2=W$ 和 $V^TA=AV$. **提示**：如果 $W=U\Lambda U^*$，其中 U 是酉矩阵，$\Lambda=\mathrm{diag}(e^{i\theta_1},\cdots,e^{i\theta_n})$，且 $0\leqslant\theta_i\leqslant 2\pi$，考虑自然平方根 $\Lambda^{1/2}\equiv\mathrm{diag}(e^{i\theta_1/2},\cdots,e^{i\theta_n/2})$，且设 $V=U\Lambda^{1/2}U^*$. 证明 $W^TA=AW$ 当且仅当 Λ 与 U^TAU 可交换. 或者利用(1.3.12)证明中的证法证明，V 是 W 的多项式，由此推出 $\Lambda^{1/2}$ 与 U^TAU 可交换，因而 $V^TA=AV$.

28. 对推论(4.4.4)的又一个证明的下述要点给出详细的论述. 记号和假设如(4.4.4)中所述. 这实质上是 Hua (1944)的证明. 首先假定 A 是非奇异的. (a)$A\bar{A}$ 是 Hermite 矩阵，且它是正定的(对所有 $x\in\mathbf{C}^n$，$x^*A\bar{A}x=(Ax)^*(\bar{A}x)\geqslant 0$)，因而存在一个酉矩阵 $Z\in M_n$ 和一个非负非奇异对角矩阵 $\Sigma\in M_n$ 使得 $A\bar{A}=Z\Sigma^2Z^*$. (b)$W\equiv\Sigma^{-1}Z^*A\bar{Z}$ 是酉矩阵且 ΣW 是对称矩阵，因而 $\Sigma W=W^T\Sigma$. (c)利用 27 题证明，存在一个酉矩阵 $V\in M_n$ 使得 $V^2=W$ 且 $\Sigma V=V^T\Sigma$. (d) $Z^*A\bar{Z}=\Sigma W=\Sigma V^2=(\Sigma V)V=V^T\Sigma V$，故 $A=(ZV^T)\Sigma(ZV^T)^T$. 设 $U=ZV^T$. (e)若 A 是奇异矩阵，利用习题 2 开头的论断把 A 化为非奇异的情形.

217

进一步阅读与注释 关于推论(4.4.4)的原型既可参看 T. Takagi, "On an Algebraic Problem Related to an Analytic Theorem of Caratheodory and Fejer and on an Allied Theorem of Landau," *Japan. J. Math.* 1 (1925), 83-93, 也可参看 I. Schur, "Ein Satz über Quadratische Formen mit Komplexen Koeffizienten," *Amer. J. Math.* 67 (1945), 472-480. 另给的几个证明可参看 C. L. Siegel, "Symplectic Geometry," *Amer. J. Math* 65 (1943), lemma 1, pp. 12, 14-15; L.-K. Hua, "On the Theory of Automorphic Functions of a Matrix Variable I-Geometric Basis," *Amer. J. Math.* 66 (1944), 470-488; 及 N. Jacobson, "Normal Semi-Linear Transformations," *Amer. J. Math.* 61 (1939), 45-58. 用三角约化(4.4.3)证明(4.4.4)可参看 Y. P. Hong and R. A. Horn, "On the Reduction of a Matrix to Triangular or Diagonal Form by Consimilarity," *SIAM J. Algebraic and Discrete Methods* (to appear). 关于推论(4.4.11)到任意域的推广可参看 O. Taussky, "The Role of Symmetric Matrices in the Study of General Matrices," *Linear Algebra Appl.* 5 (1972), 147-154.

4.5 Hermite 矩阵、对称矩阵的相合与同时对角化

任一二阶线性偏微分算子 L 可以写成形式

$$Lf=\sum_{i,j=1}^{n}a_{ij}(x)\frac{\partial^2 f(x)}{\partial x_i\,\partial x_j}+\text{诸低阶项},\quad x=[x_i]\in D\subset\mathbf{R}^n,\tag{4.5.1}$$

其中，假定系数 $a_{ij}(x)$ 定义在某个区域 $D\subset\mathbf{R}^n$ 上，函数 f 在 D 上二次连续可微. 正如在(4.0.3)所看到的那样，我们不妨假定，对所有 $x\in D$，系数矩阵 $A(x)=[a_{ij}(x)]$ 是实对称矩阵. 我们所说的低阶项指的是只包含 f 及其一阶偏导数的那些项.

如果作自变量到新变量 $s=[s_i]\in D\subset\mathbf{R}^n$ 的非奇异变换，则每个 $s_i=s[x]=s_i(x_1,\cdots,x_n)$，而非奇异性则说明 Jacobi 矩阵

$$S(x)=\left[\frac{\partial s_i(x)}{\partial x_j}\right]\in M_n$$

在 D 的每一点非奇异，这个假定保证变量 $x=x(s)$ 的逆变换局部存在. 直接应用链法则可证，在这些新坐标下，算子 L 有形式

218

$$Lf=\sum_{i,j=1}^{n}\Big[\sum_{p,q=1}^{n}\frac{\partial s_i}{\partial x_p}a_{pq}\frac{\partial s_j}{\partial x_q}\Big]\frac{\partial^2 f}{\partial s_i\,\partial s_j}+\text{诸低阶项}$$

$$=\sum_{i,j=1}^{n}b_{ij}\frac{\partial^2 f}{\partial s_i\,\partial s_j}+\text{诸低阶项}. \tag{4.5.2}$$

因此，(在坐标 $s=[s_i]$下)新的系数矩阵 B 与(在坐标 $x=[x_i]$下)旧的实系数矩阵 A 的关系可用关系式

$$B=SAS^T \tag{4.5.3^T}$$

表示，其中 S 是非奇异实矩阵.

如果微分算子 L 与某个物理定律有关(例如，Laplace 算子 $L=\nabla^2$ 和静电势)，尽管对自变量的坐标选择显然会影响 L 的形式，但它决不会影响该定律. 因此，我们不禁要问，通过关系式(4.5.3^T)与已知矩阵 A 相关联的所有矩阵 B 的集合具有什么不变量.

另一个像(4.5.3^T)那样的变换例子来源于概率与统计. 假定在有期望算子 E 的某个概率空间上，X_1，X_2，…，X_n 是具有二阶矩的实或复随机变量，且设 $\mu_i=E(X_i)$表示相应的平均值. Hermite 矩阵 $A=[a_{ij}]=(E[(X_i-\mu_i)(\bar{X}_j-\bar{\mu}_j)])\equiv\mathrm{Cov}(X)$是随机向量 $X=[X_1,\cdots,X_n]^T$ 的协方差矩阵. 如果 $S=[s_{ij}]\in M_n$ 是给定的矩阵，则 SX 是其分量为 X 的诸分量的线性组合的随机向量. SX 的诸分量的平均值是

$$E((SX)_i)=E\Big(\sum_{k=1}^{n}s_{ik}X_k\Big)=\sum_{k=1}^{n}s_{ik}E(X_k)=\sum_{k=1}^{n}s_{ik}\mu_k,$$

而 SX 的协方差矩阵是

$$\begin{aligned}\mathrm{Cov}(SX)&=(E[((SX)_i-E((SX)_i))(\overline{(SX)}_j-E(\overline{(SX)}_j))])\\&=\Big(E\Big[\Big(\sum_{p=1}^{n}s_{ip}(X_p-\mu_p)\Big)\Big(\sum_{q=1}^{n}\bar{s}_{jq}(\bar{X}_q-\bar{\mu}_q)\Big)\Big]\Big)\\&=\Big(\sum_{p,q=1}^{n}s_{ip}E[(X_p-\mu_p)(\bar{X}_q-\bar{\mu}_q)]\bar{s}_{jq}\Big)=\Big(\sum_{p,q=1}^{n}s_{ip}a_{pq}\bar{s}_{jq}\Big)\\&=SAS^*.\end{aligned}$$

这说明

$$\mathrm{Cov}(SX)=S\,\mathrm{Cov}(X)S^*. \tag{4.5.3*}$$

219

因此，随机向量的协方差矩阵的变化规律与(4.5.3^T)稍有不同. 但是，如果矩阵 S 是实的，它便简化成(4.5.3^T).

作为最后一个例子，考虑一般二次型

$$Q_A(x)=\sum_{i,j=1}^{n}a_{ij}x_ix_j=x^TAx,\quad x=[x_i]\in\mathbf{C}^n,$$

以及 Hermite 型

$$H_B(x)=\sum_{i,j=1}^{n}b_{ij}\bar{x}_ix_j=x^*Bx,\quad x=[x_i]\in\mathbf{C}^n,$$

其中 $A=[a_{ij}]$且 $B=[b_{ij}]$. 如果 $S\in M_n$ 是给定的矩阵，则

$$Q_A(Sx)=(Sx)^TA(Sx)=x^T(S^TAS)x=Q_{S^TAS}(x),$$

$$H_A(Sx) = (Sx)^* B(Sx) = x^*(S^*BS)x = H_{S^*BS}(x).$$

在这个例子中，A，B，S 和 x 为实的或为复的是无关紧要的．这里，有两种稍微不同的变换规律在起作用，而这正是给出下述定义的理由．

4.5.4 定义 设 A，$B\in M_n$ 是给定的矩阵．如果存在非奇异矩阵 S 使得

(a) $B=SAS^*$，则称 B 是*相合（"星相合"）于 A．

(b) $B=SAS^T$，则称 B 是^{T}相合（T-相合）于 A．

显然，这两个相合概念肯定有密切关系；如果 S 是实矩阵，它们是相同的．当区分这两个概念是无关紧要的时候，采用术语相合，而不加词头．有些作者用术语共轭相合表示*相合，而我们采用更便于记忆的术语．

练习 证明相合的矩阵有相同的秩．

值得提出的是，如果 A 是 Hermite 矩阵，则 SAS^* 亦是 Hermite 矩阵（即使 S 是奇异矩阵）；如果 A 是对称矩阵，则 SAS^T 也是对称矩阵．通常，对保持矩阵类型不变的相合感兴趣，例如，关于 Hermite 矩阵的*相合和关于对称矩阵的^{T}相合．但是，如果 A 是实对称矩阵，则它是对称矩阵，也是 Hermite 矩阵；于是 SAS^* 是 Hermite 矩阵，而 SAS^T 是对称矩阵．对于实对称矩阵，我们可能想按相关的内容来考虑*相合或^{T}相合．这两种相合共同具有一个重要的类似性质．

220

4.5.5 定理 *相合和^{T}相合都是等价关系．即对任一 $A\in M_n$，

(a) A 与 A 相合．

(b) 如果 A 与 B 相合，则 B 与 A 相合．

(c) 如果 A 与 B 相合且 B 与 C 相合，则 A 与 C 相合．

证明：对于(a)，我们把 A 写成 $A=IAI^*$．如果 $A=SBS^*$，且 S 是非奇异矩阵，则 $B=S^{-1}A(S^{-1})^*$．最后，如果 $A=S_1BS_1^*$ 且 $B=S_2CS_2^*$，则 $A=(S_1S_2)C(S_1S_2)^*$．对于^{T}相合的证明形式上是相同的． □

因此，所有 $n\times n$ 矩阵的集合按相合关系划分成等价类．作为一个理论问题，可以在每一相合关系下找到每个等价类的一个标准代表元．这个问题对*相合更复杂一些，所以先讨论这种情形．

通过识别相合关系的诸不变量可以辨认并划分各微分算子，这个实际问题促使我们去分析研究由（经实矩阵 S）相合于某个矩阵的诸实对称矩阵组成的等价类的标准代表元问题．结果证明，这个问题有一个简单的解答：只要计算正、负特征值和零特征值的个数．因为这个理由，引进下述专有名词．

4.5.6 定义 设 $A\in M_n$ 是 Hermite 矩阵．A 的惯性是有序三元组

$$i(A) = (i_+(A), i_-(A), i_0(A)),$$

其中，$i_+(A)$是 A 的特征值的个数，$i_-(A)$是 A 的负特征值的个数，$i_0(A)$是 A 的零特征值的个数，并且都计相重特征值的个数．注意，A 的秩等于 $i_+(A)+i_-(A)$．A 的符号差等于数值 $i_+(A)-i_-(A)$．

练习 证明，如果知道 A 的符号差和秩，则 Hermite 矩阵 $A\in M_n$ 的惯性是唯一确定的，反之亦然.

[221]

如果 $A\in M_n$ 是给定的 Hermite 矩阵，则 $A=U\Lambda U^*$，其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$ 而 U 是酉矩阵. 为方便起见，假定 Λ 的诸对角元中首先出现的是正特征值，然后是负特征值，最后是零特征值. 于是，$\lambda_1,\lambda_2,\cdots,\lambda_{i_+}>0$，$\lambda_{i_++1},\cdots,\lambda_{i_++i_-}<0$，而 $\lambda_{i_++i_-+1}=\cdots=\lambda_n=0$. 如果令

$$D=\mathrm{diag}(+\sqrt{\lambda_1},\cdots,+\sqrt{\lambda_{i_+}},+\sqrt{-\lambda_{i_++1}},\cdots,+\sqrt{-\lambda_{i_++i_-}},1,\cdots,1)$$

则 D 是非奇异实对角矩阵，且

$$\Lambda=D\begin{bmatrix}1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & 0 & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & 1 & & & \\ & & 0 & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0\end{bmatrix}D,$$

其中所展示的矩阵恰好有 $i_+(A)$ 项"$+1$"，$i_-(A)$ 项"-1"和 $i_0(A)$ 项"0". 于是，矩阵 A 可以写成

$$A=U\Lambda U^*=S\begin{bmatrix}1 & & & & & & & & \\ & \ddots & & & & & 0 & & \\ & & 1 & & & & & & \\ & & & 1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & 1 & & & \\ & & & & & & 0 & & \\ & & 0 & & & & & \ddots & \\ & & & & & & & & 0\end{bmatrix}S^*=SI(A)S^*,\quad(4.5.7)$$

其中 $S=UD$ 是非奇异矩阵，而 $I(A)$ 是 A 的惯性矩阵. 因此，每个 Hermite 矩阵*相合于一个形式很简单的对角矩阵，只要知道了该矩阵的惯性，便知道了这个对角矩阵. 用惯性矩阵作为*相合于 A 的矩阵等价类的标准代表元应当是有吸引力的，不过，要做到这一点，必须确认*相合的 Hermite 矩阵有相同的惯性. 这正是下述定理的内容，通常称该定理为 Sylveter 惯性定律.

[222]

4.5.8 定理 设 $A,B\in M_n$ 是 Hermite 矩阵，则存在非奇异矩阵 $S\in M_n$ 使得 $A=SBS^*$，当且仅当 A 和 B 有相同的惯性，即有相同个数的正、负特征值和零特征值.

证明：如果 A 和 B 有相同的惯性，则每一个矩阵都可表示成形式(4.5.7)，其中每个矩阵的 S 可能不同，但却有相同的惯性矩阵. 因为*相合关系是传递的，又 A 和 B *相合于同一个矩阵，所以它们彼此*相合，这正是要证的逆命题.

假定 A 与 B *相合，且对某个非奇异矩阵 $S\in M_n$ 有 $A=SBS^*$. 因为相合矩阵有相同的秩，所以 $i_0(A)=i_0(B)$，因而只需证明 $i_+(A)=i_+(B)$. 设 $v_1,v_2,\cdots,v_{i_+(A)}$ 是 A 的相应于正特征值 $\lambda_1(A),\cdots,\lambda_{i_+(A)}(A)$ 的正交单位向量，另外设 $S_+(A)=\mathrm{Span}\{v_1,\cdots,v_{i_+(A)}\}$.

$S_+(A)$的维数是 $i_+(A)$. 又如果 $x=\alpha_1 v_1+\cdots+\alpha_{i_+(A)}v_{i_+(A)}\neq 0$，则 $x^*Ax=\lambda_1(A)|\alpha_1|^2+\cdots+\lambda_{i_+(A)}(A)|\alpha_{i_+(A)}|^2>0$. 另一方面，

$$x^*SBS^*x=(S^*x)^*B(S^*x)>0,$$

因而，对于具有维数 $i_+(A)$的 Span$\{S^*v_1,\cdots,S^*v_{i_+(A)}\}$中所有非零向量 y 有 $y^*By>0$. 根据推论(4.3.23)，必须有 $i_+(B)\geqslant i_+(A)$. 但是，因为 A 和 B 在这个证明中的作用可以颠倒过来，所以得出 $i_+(B)=i_+(A)$. □

练习 设 $A\in M_n$ 是 Hermite 矩阵. 证明 A *相合于单位矩阵，当且仅当 A 的所有特征值都是正的.

练习 设 A，$B\in M_n$ 是对称矩阵，证明 A 和 B 可经复矩阵*相合，当且仅当它们可经实矩阵相合.

练习 设 A，$B\in M_n$ 是实对称矩阵. 证明 A 和 B 可经实矩阵相合，当且仅当 A 和 B 有相同的惯性.

练习 在由 $n\times n$ 复 Hermite 矩阵所组成的集合中，在*相合下有多少个不同的等价类？在由 $n\times n$ 实对称矩阵组成的集合中呢？

由于 Sylvester 定理保证在*相合下 Hermite 矩阵的诸特征值的符号不变，从而完全解决了在*相合下从 Hermite 矩阵的每个等价类选取一个代表元的问题. 但是在*相合下诸特征值的大小如何变化呢？利用 Weyl 定理(4.3.1)最简单的形式，可以给出 Sylvester 定理的数量形式. [223]

4.5.9 定理(Ostrowski) 设 A，$S\in M_n$，A 是 Hermite 矩阵，而 S 是非奇异矩阵. 设 A 和 SS^* 的诸特征值按递增顺序(4.2.1)排列. 则对每个 $k=1,2,\cdots$，存在正实数 θ_k，使得 $\lambda_1(SS^*)\leqslant\theta_k\leqslant\lambda_n(SS^*)$且

$$\lambda_k(SAS^*)=\theta_k\lambda_k(A). \tag{4.5.10}$$

证明：首先，若 $SS^*x=\lambda x$ 且 $x\neq 0$，则 $\lambda=x^*SS^*x/x^*x=(S^*x)^*(S^*x)/x^*x>0$，因而 SS^* 的所有特征值都是正的. 设 k 是某个整数，$1\leqslant k\leqslant n$，且考察 Hermite 矩阵 $A-\lambda_k(A)I$，它的第 k 个特征值是零. 根据 Sylvester 定理(4.5.8)，$S(A-\lambda_k(A)I)S^*=SAS^*-\lambda_k(A)SS^*$ 的第 k 个特征值也是零. Weyl 不等式(4.3.2)说明，$SAS^*-\lambda_k(A)SS^*$ 的第 k 个特征值有如下的上、下界

$$\begin{aligned}\lambda_k(SAS^*)+\lambda_1(-\lambda_k(A)SS^*)&\leqslant\lambda_k(SAS^*-\lambda_k(A)SS^*)=0\\&\leqslant\lambda_k(SAS^*)+\lambda_n(-\lambda_k(A)SS^*),\end{aligned}$$

或

$$\begin{aligned}\lambda_k(SAS^*)&\leqslant -\lambda_1(-\lambda_k(A)SS^*)=\lambda_n(\lambda_k(A)SS^*)\\&=\begin{cases}\lambda_k(A)\lambda_n(SS^*), & \text{如果 } \lambda_k(A)\geqslant 0,\\ \lambda_k(A)\lambda_1(SS^*), & \text{如果 } \lambda_k(A)\leqslant 0,\end{cases}\end{aligned}$$

且

$$\begin{aligned}\lambda_k(SAS^*)&\geqslant -\lambda_n(-\lambda_k(A)SS^*)=\lambda_1(\lambda_k(A)SS^*)\\&=\begin{cases}\lambda_k(A)\lambda_1(SS^*), & \text{如果 } \lambda_k(A)\geqslant 0,\\ \lambda_k(A)\lambda_n(SS^*), & \text{如果 } \lambda_k(A)\leqslant 0.\end{cases}\end{aligned}$$

在任何一种情形[$\lambda_k(A)\geq 0$ 或者 $\lambda_k(A)\leq 0$]下．这些不等式都推出 $\lambda_k(SAS^*)=\theta_k\lambda_k(A)$ 对适合 $\lambda_1(SS^*)\leq\theta_k\leq\lambda_n(SS^*)$ 的某个 θ_k 成立． □

在 Ostrowki 定理中，如果 $A=I\in M_n$，则所有 $\lambda_k(A)=1$ 且 $\theta_k=\lambda_k(SS^*)$．如果 $S\in M_n$ 是酉矩阵，则 $\lambda_1(SS^*)=\lambda_n(SS^*)=1$ 且所有 $\theta_k=1$；这表明在酉相似下特征值的不变性．因此，定理中给出的关于 θ_k 的界对任一给定的 A 和任一给定的非奇异矩阵 S 是最合适的．

通过简单的连续性论证，Ostrowski 定理可以推广到包括 S 是奇异矩阵的情形．在这种情形，设 $\varepsilon>0$，然后用 $S+\varepsilon I$ 代替 S 来应用定理可知 $\lambda_k((S+\varepsilon I)A(S+\varepsilon I)^*)=\theta_k\lambda_k(A)$，并且 $\lambda_1((S+\varepsilon I)(S+\varepsilon I)^*)\leq\theta_k\leq\lambda_n((S+\varepsilon I)(S+\varepsilon I)^*)$．现在让 $\varepsilon\to 0$ 便得到界 $0\leq\theta_k\leq\lambda_n(SS^*)$．这个结果可以看作 Sylvester 惯性定律到奇异*相合的推广． 224

4.5.11 推论 设 $A, S\in M_n$，且 A 是 Hermite 矩阵．设 A 和 SS^* 的诸特征值按递增顺序(4.2.1)排列．那么对每个 $k=1, 2, \cdots, n$，存在非负实数 θ_k 使得 $\lambda_1(SS^*)\leq\theta_k\leq\lambda_n(SS^*)$ 且

$$\lambda_k(SAS^*)=\theta_k\lambda_k(A).$$

特别是 SAS^* 的正(负)特征值的个数小于或等于 A 的正(负)特征值的个数．

求复对称矩阵在^{T}相合下的诸等价类的标准代表元问题有一个更简便的解法：只要计算秩．

4.5.12 定理 设 $A, B\in M_n$ 是(复或实)对称矩阵．那么，存在非奇异矩阵 $S\in M_n$ 使得 $A=SBS^T$，当且仅当 A 和 B 有相同的秩．

证明：如果 $A=SBS^T$，且 S 非奇异，则由(0.4.6)可知，A 与 B 有相同的秩．反过来，利用(4.4.4)可导出

$$A=U_1\Sigma_1U_1^T=U_1I(\Sigma_1)D_1^2U_1^T=(U_1D_1)I(\Sigma_1)(U_1D_1)^T,$$

其中，$I(\Sigma_1)$ 是 Σ_1 的惯性矩阵(4.5.7)，它由 A 的秩完全确定，U_1 是酉矩阵，$\Sigma_1=\mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$ 且所有 $\sigma_i\geq 0$，$D_1=\mathrm{diag}(d_1, d_2, \cdots, d_n)$ 且

$$d_i=\begin{cases}\sqrt{\sigma_i}, & \text{如果 } \sigma_i>0,\\ 1, & \text{如果 } \sigma_i=0.\end{cases}$$

注意，D_1 是非奇异矩阵．也可以用同样的方式导出 $B=(U_2D_2)I(\Sigma_2)(U_2D_2)^T$，且其中各矩阵有类似的定义．如果假定 rank A=rank B，则 $I(\Sigma_1)=I(\Sigma_2)$，且

$$I(\Sigma_1)=(U_1D_1)^{-1}A[(U_1D_1)^T]^{-1}=I(\Sigma_2)=(U_2D_2)^{-1}B[(U_2D_2)^T]^{-1},$$

因此

$$A=(U_1D_1)(U_2D_2)^{-1}B[(U_1D_1)(U_2D_2)^{-1}]^T.$$

由此得出 A 与 B ^{T}相合． □

练习 在^{T}相合下，$n\times n$ 复对称矩阵组成的集合中有多少不同的等价类？在 $n\times n$ 实对称矩阵组成的集合中呢？ 225

练习 设 $A\in M_n$ 是对称矩阵．证明存在非奇异矩阵 $S\in M_n$ 使得 $A=SS^T$，当且仅当 A 是非奇异矩阵．

练习 设 $A, B\in M_n$ 是对称矩阵．证明存在非奇异矩阵 $X, Y\in M_n$ 使得 $A=XBY$，也就是说，A 与 B 等价，当且仅当存在非奇异矩阵 $S\in M_n$ 使得 $A=SBS^T$，即 A 与 B ^{T}相合．**提示**：

如果 $A=ZBY$，A 和 B 的秩是什么关系?

上述结果相当于关于复矩阵的 ^{T} 相合的 Sylvester 惯性定律. 下述结果相当于 Sylvester 定理的 Ostrowski 数量形式[(4.5.9)和(4.5.11)].

4.5.13 定理　设 A，$S\in M_n$，且 $A=A^T$. 设 $A=U\Sigma U^T$ 和 $SAS^T=VMV^T$ 是 A 和 SAS^T 的 Takagi 分解(4.4.4)，其中 U 和 V 是酉矩阵，$\Sigma=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_n)$，$M=\mathrm{diag}(\mu_1,\mu_2,\cdots,\mu_n)$，且所有 σ_i，$\mu_i\geqslant 0$. 设 $\lambda_i(SS^*)$ 表示 SS^* 的特征值. 假定数 σ_i，μ_i 和 $\lambda_i(SS^*)$ 都按递增顺序(4.2.1)排列. 则对每个 $k=1,2,\cdots,n$，存在适合 $\lambda_1(SS^*)\leqslant\theta_k\leqslant\lambda_n(SS^*)$ 的非负实数 θ_k 使得 $\mu_k=\theta_k\sigma_k$. 如果 S 非奇异，则所有 $\theta_i>0$.

证明：数 μ_k^2 是 BB^* 的特征值，其中 $B=SAS^T$. 因而

$$\mu_k^2=\lambda_k(BB^*)=\lambda_k(SAS^T\bar{S}\bar{A}S^*)=\lambda_k(S[AS^T\bar{S}\bar{A}]S^*)=\hat{\theta}_k\lambda_k(AS^T\bar{S}\bar{A})$$

对适合 $\lambda_1(SS^*)\leqslant\hat{\theta}_k\leqslant\lambda_n(SS^*)$ 的某个 $\hat{\theta}_k$ 成立；为得到最后一个等式，我们利用了(4.5.11). 因为两个矩阵乘积的特征值与乘积(1.3.20)的顺序无关，又因为特征值 λ_k 是实数，所以还有

$$\mu_k^2=\hat{\theta}_k\lambda_k(AS^T\bar{S}\bar{A})=\hat{\theta}_k\lambda_k(\bar{S}\bar{A}AS^T)=\hat{\theta}_k\lambda_k(SA\bar{A}S^*).$$

再应用(4.5.11)，则对适合 $\lambda_1(SS^*)\leqslant\tilde{\theta}_k\leqslant\lambda_n(SS^*)$ 的某个 $\tilde{\theta}_k$ 有

$$\mu_k^2=\hat{\theta}_k\tilde{\theta}_k\lambda_k(A\bar{A})=\hat{\theta}_k\tilde{\theta}_k\sigma_k^2.$$

因此，$\mu_k=\sqrt{\hat{\theta}_k\tilde{\theta}_k}\,\sigma_k=\theta_k\sigma_k$，且 $\theta_k=\sqrt{\hat{\theta}_k\tilde{\theta}_k}$ 在所要求的上下界之间. □

我们从(1.3.19)得知，两个可对角化的矩阵可经同一个相似变换同时对角化，当且仅当它们可交换. 关于通过相合同时对角化的相应结果是什么呢?

或许最早是由于研究关于稳定平衡的"最小振动"力学问题，才促使人们去考虑关于通过相合同时对角化的有关结果. 如果动力系统的组态由广义(Lagrange)坐标 $q_1,q_2,\cdots,q_n$ 来确定，其中原点是稳定平衡点，则在原点附近，势能函数 V 可以通过用广义坐标 q_i 表示的实二次型

226

$$V=\sum_{i,j=1}^{n}a_{ij}q_iq_j$$

来逼近. 动能 T 可以通过用广义速度 $\dot{q}_i$ 表示的实二次型

$$T=\sum_{i,j=1}^{n}b_{ij}\dot{q}_i\dot{q}_j$$

来逼近. 系统的变化过程由 Lagrange 方程组

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial\dot{q}_i}\right)-\frac{\partial T}{\partial q_i}+\frac{\partial V}{\partial q_i}=0$$

所决定. 它是常系数二阶线性常微分方程组，如果两个二次型 T 和 V 是非对角的，这些方程就是耦合的(因而要解这些方程是较困难的). 我们可以假定实矩阵 $A=[a_{ij}]$ 和 $B=[b_{ij}]$ 是对称的.

如果可以求得非奇异变换 $S=[s_{ij}]\in M_n$ 使得 SAS^T 和 SBS^T 都是对角矩阵，则关于适合关系

$$q_i=\sum_{j=1}^{n}s_{ij}p_j \tag{4.5.14}$$

的新广义坐标 p_i，动能二次型 T 和势能二次型 V 都是对角矩阵. 在这种情形，Lagrange 方程组就是由 n 个分离的常系数二阶线性常微分方程组成的非耦合组. 利用指数函数和三角函数不

难解出这些方程，而原问题的解可利用(4.5.14)求得.

因此，如果可以通过相合同时对角化两个实对称矩阵，则一类重要的力学问题的实质性简化是可以实现的. 根据物理知识，动能二次型是正定的，结果证明，这是可通过相合同时对角化的充分(而不是必要的)条件.

[227] 我们可能要考虑的同时对角化结果有多种形式. 可能有两个 Hermite 矩阵 A 和 B，并且可能希望对某个酉矩阵 U 使 UAU^* 和 UBU^* 都是对角矩阵，或者可能满足于较弱的结果，对某个非奇异矩阵 S 使 SAS^* 和 SBS^* 都是对角矩阵. 类似地，如果 A 和 B 是对称矩阵，我们可能希望 UAU^T 和 UBU^T，或 SAS^T 和 SBS^T 都是对角矩阵. 甚至可能会有这样的混合问题，A 是 Hermite 矩阵，而 B 是对称矩阵，希望 UAU^* 和 UBU^T，或 SAS^* 和 SBS^T 都是对角矩阵. 在每种情形，要考虑的自然相合是保持相应矩阵的特殊代数特征的相合. 所有这些情形都出现在应用之中. 它们都可以用同样的技巧来处理，而要考虑的最简单情形是两个矩阵之中有一个是非奇异的情形. 在表 4.5.15T 中列出了若干结果，它对每种情形都给出一系列等价的必要充分条件. 将这些必要充分条件按指定顺序编号是为了显示各种情形中相平行的条件.

4.5.15 定理 设 A，$B\in M_n$ 是给定的. 设 U 表示酉矩阵，S 表示非奇异矩阵，且 U，$S\in M_n$，则有表 4.5.15T.

表 4.5.15T

对 A 和 B 的假定	已对角化矩阵偶	关于可同时对角化的等价的必要充分条件
Ⅰ. $A=A^*$ $B=B^*$ A 是非奇异矩阵 $C=A^{-1}B$	(a)UAU^* 和 UBU^*	(1)存在酉矩阵 $V\in M_n$ 使得 V^*CV 是实对角矩阵 (2)C 有实特征值且可酉对角化 (3)C 是 Hermite 矩阵 (4)$AB=BA$，即 A 与 B 可交换
	(b)SAS^* 和 SBS^*	(1)存在非奇异矩阵 $R\in M_n$ 使得 $R^{-1}CR$ 是实对角矩阵 (2)C 有实特征值且可对角化
Ⅱ. $A=A^T$ $B=B^T$ A 是非奇异矩阵 $C=A^{-1}B$	(a)UAU^T 和 UBU^T	(1)存在酉阵 $V\in M_n$ 使得 V^*CV 是对角矩阵 (2)C 是可酉对角化矩阵 (3)C 是正规矩阵
	(b)SAS^T 和 SBS^T	(1)存在非奇异矩阵 $R\in M_n$ 使得 $R^{-1}CR$ 是对角矩阵 (2)C 是可对角化矩阵
Ⅲ. $A=A^*$ $B=B^T$ 如果 A 是非奇异矩阵， 令 $C=A^{-1}B$. 如果 B 是非奇异矩阵， 令 $C=B^{-1}A$	(a)UAU^* 和 UBU^T	(1)存在酉矩阵 $W\in M_n$ 使得 $W^{-1}C\bar{W}$ 是对角矩阵 (3)C 是对称矩阵 (4)$AB=BA$
	(b)SAS^* 和 SBS^T	(1)存在非奇异矩阵 $R\in M_n$ 使得 $R^{-1}C\bar{R}$ 是对角矩阵 (5)存在非奇异矩阵 $R\in M_n$ 使得 $R^{-1}C\bar{R}$ 是对称矩阵

证明： 在六组条件的每一组中，所述大部分条件的等价性只是属于定义的推论. Ⅰ(a)组中(3)和(4)的等价性可以从以下论断推出：如果 A 和 B 是 Hermite 矩阵，则 AB 是 Hermite

矩阵当且仅当 A 与 B 可交换，以及 A 是 Hermite 矩阵当且仅当 A^{-1}是 Hermite 矩阵. 类似地可证明Ⅲ(a)(3)和(4)的等价性，这是因为，B 是对称矩阵当且仅当 B^{-1}是对称矩阵，还因为，如果 A 是 Hermite 矩阵，则 $A^T=\bar{A}$.

在六组条件的每一组中，条件(1)的必要性可直接从相应的相合具有对角形式的假设推出. 例如，在情形Ⅱ(b)，如果 $SAS^T=\Lambda$ 和 $SBS^T=M$ 都是对角矩阵，则

$$A^{-1}B=(S^T\Lambda^{-1}S)[S^{-1}M(S^T)^{-1}]=S^T(\Lambda^{-1}M)(S^T)^{-1},$$

因而 $R=S^T$ 将对角化 $C=A^{-1}B$. 类似地，在情形Ⅰ(b)和Ⅲ(b)，$R=S^*$ 也将起对角化作用. 如果 S 是酉矩阵，则每种情形的相应矩阵 R 也是酉矩阵.

考虑情形Ⅰ，其中 A 和 B 是 Hermite 矩阵，且 A 是非奇异矩阵. 假定Ⅰ(b)(1)成立，即存在非奇异矩阵 $R=[r_1r_2\cdots r_n]\in M_n$，每个 $r_i\in\mathbf{C}^n$，以及对角矩阵 $\Lambda=\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，其中所有 λ_i 是实数，使得 $R^{-1}A^{-1}BR=\Lambda$，因而 $BR=AR\Lambda$ 且 $R^*BR=R^*AR\Lambda$. 不失一般性，假定相重的 λ_i 项的值是排放在一起的，因而 Λ 有分块形式

228～229

$$\Lambda=\begin{bmatrix}\Lambda_1 & & & 0\\ & \Lambda_2 & & \\ 0 & & \ddots & \\ & & & \Lambda_k\end{bmatrix},\tag{4.5.16}$$

$$\Lambda_i\in M_{n_i};\quad 1\leqslant n_i\leqslant n;\quad \Lambda_i=\mu_iI,\quad i=1,2,\cdots,k,$$

其中，所有 μ_i 是实数，且如果 $i\neq j$，则 $\mu_i\neq\mu_j$. 如果所有 λ_i 项未必相等，选取适合 $1\leqslant i$，$j\leqslant n$ 的任意 i，j 使 $\lambda_i\neq\lambda_j$，并且考察恒等式 $R^*BR=R^*AR\Lambda$ 两边的 i，j 项. 这就是

$$r_i^*Ar_j\lambda_j=r_i^*Br_j=\overline{r_j^*Br_i}=\overline{r_j^*Ar_i\lambda_i}=r_i^*Ar_j\lambda_i,$$

这里，用到了 A 和 B 是 Hermite 矩阵(因而对所有 x，$y\in\mathbf{C}^n$ 有 $x^*Ay=\overline{y^*Ax}$)以及 λ_i 和 λ_j 是实数的事实. 因为 $\lambda_i\neq\lambda_j$，我们推出 $r_i^*Ar_j=0$ 因而 $r_j^*Ar_i=r_i^*Br_j=r_j^*Br_i=0$. 这表明矩阵 R^*BR 和R^*AR 都是分块对角矩阵，且与(4.5.16)有相同的形式；即

$$R^*BR=\begin{bmatrix}B_1 & & & 0\\ & B_2 & & \\ 0 & & \ddots & \\ & & & B_k\end{bmatrix}=R^*AR\Lambda$$

$$=\begin{bmatrix}\mu_1A_1 & & & 0\\ & \mu_2A_2 & & \\ 0 & & \ddots & \\ & & & \mu_kA_k\end{bmatrix},$$

其中，对 $i=1,2,\cdots,k$，B_i，$A_i\in M_{n_i}$. 这部分化简到对角形式，如果 $k=n$，即如果所有 λ_i 都互不相同，它将完全化简成了对角形式. 如果 $k<n$，则某个子块有 $n_i>1$，且 $B_i=\mu_iA_i$. 因为 A_i 和 B_i 是 Hermite 矩阵，可以利用谱定理(4.1.5)导出 $A_i=U_iD_iU_i^*$，其中 U_i，$D_i\in M_{n_i}$，U_i 是酉矩阵，而 D_i 是实对角矩阵，则 $B_i=\mu_iA_i=U_i(\mu_iD_i)U_i^*$ 也可对角化. 如果令

$$U=\begin{bmatrix}U_1 & & & 0\\ & U_2 & & \\ 0 & & \ddots & \\ & & & U_k\end{bmatrix},\quad D=\begin{bmatrix}D_1 & & & 0\\ & D_2 & & \\ 0 & & \ddots & \\ & & & D_k\end{bmatrix},$$

且当 $n_i=1$ 时 $U_i=[1]$，则 U 是酉矩阵，D 是实对角矩阵，且 [230]

$$R^* BR=U(D\Lambda)U^*,\quad R^* AR=UDU^*.$$

最后，欲求的表示式是

$$A=[(R^{-1})^* U]D[(R^{-1})^* U]^* \text{ 和 } B=[(R^{-1})^* U](D\Lambda)[(R^{-1})^* U]^*.$$

注意，如果假定Ⅰ(a)(1)成立，证法是相同的，只是我们还知道 R 是酉矩阵. 在这种情形，$(R^{-1})^* U=RU$ 是酉矩阵且Ⅰ(a)(1)的充分性得证.

余下的四种情形中所要作的证明是类似的. 利用相应的假设得到相合矩阵，它们是分块对角矩阵，然后利用关于 Hermite 矩阵的谱定理或关于对称矩阵的 Takagi 分解(4.4.4)便完成了到对角形式的化简.

考虑情形Ⅱ，其中，A 和 B 是对称矩阵，且 A 是非奇异矩阵. 假定Ⅱ(b)(1)成立，即存在非奇异矩阵 $R=[r_1 r_2\cdots r_n]\in M_n$，每个 $r_i\in\mathbf{C}^n$，以及(不一定是实的)对角矩阵 $\Lambda=\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，使得 $R^{-1}A^{-1}BR=\Lambda$，因而 $BR=AR\Lambda$ 且 $R^TBR=R^TAR\Lambda$. 又假定相重的 λ_i 项是排放在一起的，因而 Λ 有形式(4.5.16)，且所有 μ_i 互不相同. 如果不是所有 λ_i 都相等，选取适合 $1\leqslant i,\ j\leqslant n$ 的任意 i，j 使 $\lambda_i\neq\lambda_j$，并考察恒等式 $R^TBR=R^TAR\Lambda$ 两边的 i，j 元. 这就是

$$r_i^TAr_j\lambda_j=r_i^TBr_j=r_j^TBr_i=r_j^TAr_i\lambda_i=r_i^TAr_j\lambda_i$$

这里，用到了 A 和 B 的对称性(对所有 $x,\ y\in\mathbf{C}^n$，$x^TAy=y^TAx$). 因 $\lambda_i\neq\lambda_j$，推出 $r_i^TAr_j=0$，因而 $r_j^TAr_i=r_j^TBr_i=r_i^TBr_j=0$. 这表明矩阵 R^TBR 和 R^TAR 都是分块对角矩阵且与(4.5.16)有相同的形式；即

$$R^TBR=\begin{bmatrix}B_1 & & & 0\\ & B_2 & & \\ 0 & & \ddots & \\ & & & B_k\end{bmatrix}=R^TAR\Lambda$$

$$=\begin{bmatrix}\mu_1A_1 & & & 0\\ & \mu_2A_2 & & \\ 0 & & \ddots & \\ & & & \mu_kA_k\end{bmatrix},$$

其中 B_i，$A_i\in M_{n_i}$. 如果 $k=n$，这就是所要求的化简. 如果 $k<n$，则某个子块有 $n_i>1$ 且 $B_i=\mu_iA_i$. 因为 A_i 和 B_i 都是对称矩阵. 可以利用 Takagi 分解(4.4.4)导出 $A_i=U_i\Sigma_iU_i^T$，其中，U_i，$\Sigma_i\in M_{n_i}$，U_i 是酉矩阵，而 Σ 是具有非负对角元的对角矩阵. 于是 $B_i=\mu_iA_i=U_i(\mu_i\Sigma_i)U_i^T$. 如果令 [231]

$$U=\begin{bmatrix}U_1 & & & 0\\ & U_2 & & \\ 0 & & \ddots & \\ & & & U_k\end{bmatrix},\quad \Sigma=\begin{bmatrix}\Sigma_1 & & & 0\\ & \Sigma_2 & & \\ 0 & & \ddots & \\ & & & \Sigma_k\end{bmatrix},$$

且当 $n_i=1$ 时 $U_i=[1]$，则 U 是酉矩阵，Σ 是(具有非负对角元的)对角矩阵，且

$$R^TBR=U(\Sigma\Lambda)U^T \text{ 和 } R^TAR=U\Sigma U^T.$$

最后，欲求的表示式是

$$A=[(R^{-1})^TU]\Sigma[(R^{-1})^TU]^T \text{ 和 } B=[(R^{-1})^TU]\Sigma\Lambda[(R^{-1})^TU]^T.$$

如果假定Ⅱ(a)(1)成立，则 R 是酉矩阵且 $(R^{-1})^TU=\bar{R}U$ 是酉矩阵，因此Ⅱ(a)(1)的充分性也得到了证明.

在情形Ⅲ，证明需作一点修改. 设 $A, B\in M_n$，且 A 是非奇异的 Hermite 矩阵，B 是对称矩阵. 假定Ⅲ(b)(1)成立，即存在非奇异矩阵 $R=[r_1r_2\cdots r_n]\in M_n$ 和对角矩阵 $\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 使得 $R^{-1}A^{-1}B\bar{R}=\Lambda$. 因而 $B\bar{R}=AR\Lambda$ 且 $R^*B\bar{R}=\bar{R}^TB\bar{R}=R^*AR\Lambda$. 现在假定模相同的 λ_i 项排放在一起使 Λ 有形式

$$\Lambda=\begin{bmatrix}\Lambda_1 & & & 0\\ & \Lambda_2 & & \\ 0 & & \ddots & \\ & & & \Lambda_k\end{bmatrix},$$

$$\text{其中},\Lambda_i=\begin{bmatrix}\mu_i^{(1)} & & & 0\\ & \mu_i^{(2)} & & \\ 0 & & \ddots & \\ & & & \mu_i^{(n_i)}\end{bmatrix},\quad i=1,\cdots,k,$$

并且对 $j, k=1, 2, \cdots, n_i$ 有 $|\mu_i^{(j)}|=|\mu_i^{(k)}|$，而如果 $i\neq j$，则 $|\mu_i^{(p)}|\neq|\mu_j^{(q)}|$. 如果不是所有 λ_i 项有相同的模，则选取适合 $1\leqslant i, j\leqslant n$ 和 $|\lambda_i|\neq|\lambda_j|$ 的任意 i, j，然后考察恒等式 $R^TBR=R^*AR\Lambda$ 两边的 i, j 元. 这就是

[232]

$$r_i^*Ar_j\lambda_j=\bar{r}_i^TB\bar{r}_j=\bar{r}_j^TB\bar{r}_i=r_j^*Ar_i\lambda_i=\overline{r_i^*Ar_j}\lambda_i,$$

这里，用到了 A 是 Hermite 矩阵及 B 是对称矩阵的事实. 于是 $|r_i^*Ar_j||\lambda_j|=|r_i^*Ar_j||\lambda_i|$. 又因为 $|\lambda_i|\neq|\lambda_j|$，由此推出 $r_i^*Ar_j=0$，因而 $r_j^*Ar_i=\bar{r}_i^TB\bar{r}_j=\bar{r}_j^TBr_i=0$. 这表明矩阵 $\bar{R}^TB\bar{R}$ 和 R^*AR 都是分块对角矩阵且与(4.5.16)有相同的形式，即

$$\bar{R}^TB\bar{R}=\begin{bmatrix}B_1 & & & 0\\ & B_2 & & \\ 0 & & \ddots & \\ & & & B_k\end{bmatrix}=R^*AR\Lambda,$$

$$\begin{bmatrix} A_1\Lambda_1 & & & 0 \\ & A_2\Lambda_2 & & \\ & & \ddots & \\ 0 & & & A_k\Lambda_k \end{bmatrix},$$

其中，所有 B_i，A_i，$\Lambda_i \in M_{n_i}$，且 $\Lambda_i = \sigma_i D_i^2$，$\sigma_i \geqslant 0$，

$$D_j = \mathrm{diag}(e^{i\theta_1}, e^{i\theta_2}, \cdots, e^{i\theta_{n_j}}),$$

所有 $\theta_j \in \mathbf{R}$. 如果 $k=n$，这就是所要求的化简. 如果 $k<n$，那么某个子块有 $n_i>1$ 且 $B_i = A_i\Lambda_i = \sigma_i A_i D_i^2$. 因为 D_i 是酉对角矩阵，所以 $D_i^* = \bar{D}_i = \bar{D}_i^T = D_i^{-1}$，因而

$$\bar{D}_i^T B_i \bar{D}_i = \sigma_i D_i^* A_i D_i. \tag{4.5.17}$$

这个恒等式左边是对称矩阵 $D_i^T B_i D_i$，而右边是 Hermite 矩阵 $\sigma_i D_i^* A_i D_i$，且所有 σ_i 是实数. 如果 $\sigma_i \neq 0$，推出 $D_i^* A_i D_i$ 是 Hermite 矩阵，又是对称矩阵. 而一个 Hermite 矩阵只有在它是实矩阵时才能是对称矩阵. 因而，如果 $\sigma_i \neq 0$，$D_i^* A_i D_i$ 就是实对称矩阵. 如果 $\sigma_i = 0$（至多对 i 的一个值可能出现这种情形），则 $D_i^* A_i D_i$ 是 Hermite 矩阵，但不一定是实矩阵. 根据谱定理，对每个 $i=1,\cdots,k$ 存在酉矩阵 $U_i \in M_{n_i}$ 以及实对角矩阵 M_i，使得 $D_i^* A_i D_i = U_i M_i U_i^*$. 如果 $\sigma_i \neq 0$，则 U_i 可以选为实正交矩阵，这时 $U_i^T = U_i^*$ 且

$$\bar{D}_i^T B_i \bar{D}_i = \sigma_i D_i^* A_i D_i = U_i(\sigma_i M_i)U_i^T.$$

如果 $\sigma_i = 0$，则 $U_i^* = U_i^T$ 可能不成立，但是，因为两边都是零，所给出的等式仍然正确. 因此，对所有 $i=1,2,\cdots,k$，有

$$A_i = (D_i U_i) M_i (D_i U_i)^* \text{ 和 } B_i = (D_i U_i)(\sigma_i M_i)(D_i U_i)^T.$$

如果令 233

$$U = \begin{bmatrix} D_1U_1 & & & 0 \\ & D_2U_2 & & \\ & & \ddots & \\ 0 & & & D_kU_k \end{bmatrix},$$

$$M = \begin{bmatrix} M_1 & & & 0 \\ & M_2 & & \\ & & \ddots & \\ 0 & & & M_k \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1 I & & & 0 \\ & \sigma_2 I & & \\ & & \ddots & \\ 0 & & & \sigma_k I \end{bmatrix}.$$

则 $B = [(\bar{R}^{-1})^T U]\Sigma M[U^T \bar{R}^{-1}]$ 且 $A = [(R^{-1})^* U]M[U^* R]$，这正是所要证明的. 如果假定Ⅲ(a)(1)成立，则 R 是酉矩阵，因而 $(R^{-1})^* U = RU$ 和 $(\bar{R}^{-1})^T U = RU$ 是酉矩阵，因此Ⅲ(a)(1)的充分性得证.

当 A 是非奇异矩阵时，这就完成了Ⅲ的证明. 如果 B 是非奇异矩阵，Ⅲ(b)(1)的假定说明，存在非奇异矩阵 $R \in M_n$ 使得 $R^{-1}B^{-1}AR = \Lambda$ 是对角矩阵，因而 $AR = BR\Lambda$ 且 $\bar{R}^* A\bar{R} = R^* BR\Lambda$. 此后的证明形式上与 A 是非奇异矩阵的情形相同. 在证明中仅仅交换 A 和 B 的地

位，且用 Takagi 分解(4.4.4)对角化 $D_i^T B_i D_i$，而不是用谱定理. □

在定理(4.5.15)(表 4.5.15T)的情形Ⅰ和Ⅱ，有一个关于 $A^{-1}B$ 的熟知的条件，它等价于 A 和 B 可通过相应的相合同时对角化. 这就是 $A^{-1}B$ 可对角化(或许其特征值都是实的)，即 $A^{-1}B$ 具有形式 $R\Lambda R^{-1}$，其中 Λ 是对角矩阵(或许 Λ 是实矩阵). 原则上，这个条件可以通过验证 $A^{-1}B$ 的极小多项式是否有不同的线性(或许是实的)因式来检验. 但是在情形Ⅲ，所述条件是不多见的，即 $A^{-1}B$ 具有形式 $R\Lambda\bar{R}^{-1}$，其中 Λ 是对角矩阵. 这个条件说明 $A^{-1}B$ 可通过合相似而不是通常的相似对角化. 关于合相似性的讨论见(4.6)节. 定理(4.6.11)证明，条件(4.5.15 Ⅲ(b)(1))等价于条件：$C\bar{C}$ 的特征值均为非负实数，$C\bar{C}$ 可对角化，且 $\operatorname{rank} C=\operatorname{rank} C\bar{C}$.

在定理(4.5.15)中作非奇异性假定是方便的. 但是在酉相合的情形Ⅰ(a)、Ⅱ(a)和Ⅲ(a)中，这个假定可以取消. 在情形Ⅰ(a)，这种计算方法给出了关于可交换的 Hermite 矩阵可同时酉对角化的经典结果(4.1.6)的又一个证明. [234]

4.5.18 推论 设 $A, B\in M_n$.

(a) 如果 A 和 B 都是 Hermite 矩阵，则存在酉矩阵 $U\in M_n$ 使得 UAU^* 和 UBU^* 都是对角矩阵，当且仅当 AB 是 Hermite 矩阵：即 $AB=BA$.

(b) 如果 A 和 B 都是对称矩阵，则存在酉矩阵 $U\in M_n$ 使得 UAU^T 和 UBU^T 都是对角矩阵，当且仅当 $A\bar{B}$ 是正规矩阵：即 $A\bar{B}BA=B\bar{A}A\bar{B}$.

(c) 如果 A 是 Hermite 矩阵，而 B 是对称矩阵，则存在酉矩阵 $U\in M_n$，使得 UAU^* 和 UBU^T 都是对角矩阵，当且仅当 AB 是对称矩阵：即 $AB=B\bar{A}$.

证明：(a) 如果 $UAU^*=\Lambda$ 和 $UBU^*=M$ 都是对角矩阵，则 $A=U^*\Lambda U$，$B=U^*MU$，因而 $AB=U^*\Lambda UU^*MU=U^*\Lambda MU=U^*M\Lambda U=U^*MUU^*\Lambda U=BA$. 反过来，如果 $AB=BA$，则对某个 $\varepsilon>0$，$A_\varepsilon=A+\varepsilon I$ 是非奇异 Hermite 矩阵，且 $A_\varepsilon B=(A+\varepsilon I)B=AB+\varepsilon B=BA+\varepsilon B=B(A+\varepsilon I)=BA_\varepsilon$. 因此，$B$ 与 A_ε 和 A_ε^{-1} 可交换，因而 $A_\varepsilon^{-1}B$ 是 Hermite 矩阵根据(4.5.15)(表 4.5.15T)的Ⅰ(a)(3)，存在酉矩阵 U_ε，使得 $U_\varepsilon A_\varepsilon U_\varepsilon^*=U_\varepsilon AU_\varepsilon^*+\varepsilon I=\Lambda_\varepsilon$ 和 $U_\varepsilon BU_\varepsilon^*=M_\varepsilon$ 都是对角矩阵，因而 $U_\varepsilon AU_\varepsilon^*=\Lambda_\varepsilon-\varepsilon I$ 和 $U_\varepsilon BU_\varepsilon^*=M_\varepsilon$ 都是对角矩阵.

(b) 如果 $UAU^T=\Lambda$ 和 $UBU^T=M$ 都是对角矩阵，则 $A=U^*\Lambda\bar{U}$，$B=U^*M\bar{U}$，且 $AB=U^*\Lambda\bar{U}U^T\bar{M}U=U^*(\Lambda\bar{M})U$ 可酉对角化，因而是正规矩阵. 关于逆命题，假定 $A\bar{B}$ 是正规矩阵且 A 是非奇异矩阵，则 $A\bar{B}=(A^{-1})^{-1}\bar{B}$ 是正规矩阵，而(4.5.15)的Ⅱ(a)(3)说明，两个对角矩阵 A^{-1} 和 B 是同时可酉对角化的. 因此，存在酉矩阵 $U\in M_n$ 和对角矩阵 Λ，$M\in M_n$，使得 $A^{-1}=U\Lambda U^T$ 和 $\bar{B}=UMU^T$. 于是 $A=\bar{U}\Lambda^{-1}U^*$ 和 $B=\bar{U}\bar{M}\bar{U}^T$，这正是所要求的 A 和 B 同时对角化形式. 如果 A 是奇异矩阵，则根据(4.4.4)，存在酉矩阵 $U\in M_n$，使得 UAU^T 是对角矩阵，如果必要，还可以交换 U 的诸列使得

$$UAU^T=\begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix},\quad \Sigma\in M_k,\quad 1\leqslant k<n.$$

且 Σ 是非奇异对称矩阵(实际上是对角矩阵). 如果把 UBU^T 写成相应的分块形式

$$UBU^T=\begin{bmatrix}B_{11} & B_{12}\\ B_{12}^T & B_{22}\end{bmatrix},\quad B_{11}\in M_k,\quad B_{22}\in M_{n-k},$$

则子块 B_{11} 和 B_{22} 是对称矩阵，且有

$$(UAU^T)(\overline{UBU^T}) = UA\bar{B}U^* = \begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\bar{B}_{11} & \bar{B}_{12}\\ \bar{B}_{12}^T & \bar{B}_{22}\end{bmatrix} = \begin{bmatrix}\Sigma\bar{B}_{11} & \Sigma\bar{B}_{12}\\ 0 & 0\end{bmatrix}.$$

235

但是 $UA\bar{B}U^*$ 还是正规矩阵，因而 $\Sigma\bar{B}_{12}=0$（见本节末习题 20），又因为 Σ 是非奇异矩阵，所以 $B_{12}=0$. 这说明

$$UAU^T = \begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix},\quad UBU^T = \begin{bmatrix}B_{11} & 0\\ 0 & B_{22}\end{bmatrix},$$

且

$$(UAU^T)(\overline{UBU^T}) = \begin{bmatrix}\Sigma\bar{B}_{11} & 0\\ 0 & 0\end{bmatrix}.$$

根据上述关于非奇异情形的证明，得知，存在酉矩阵 $V_1\in M_k$ 和对角矩阵 Λ_1，$\Lambda_2\in M_k$，使得 $\Sigma=V_1\Lambda_1V_1^T$ 和 $B_{11}=V_1\Lambda_2V_1^T$. 因为 B_{22} 是对称矩阵，我们还知道，存在酉矩阵 $V_2\in M_{n-k}$ 和对角矩阵 $\Lambda_3\in M_{n-k}$，使得 $B_{22}=V_2\Lambda_3V_2^T$. 如果设 $\Lambda=\Lambda_1\oplus 0\in M_n$，$M=\Lambda_2\oplus\Lambda_3$，且 $V=V_1\oplus V_2$，则有 $UAU^T=V\Lambda V^T$，$UBU^T=VMV^T$. 因此，$A=(U^*V)\Lambda(U^*V)^T$ 和 $B=(U^*V)M(U^*V)^T$ 是所有要求的同时对角化形式.

(c) 如果 $UAU^*=\Lambda$ 和 $UBU^T=M$ 都是对角矩阵，则 Λ 一定是实矩阵. 有 $A=U^*\Lambda U$，$B=U^*M\bar{U}$ 以及

$$\begin{aligned}A\bar{B} &= U^*\Lambda UU^T\bar{M}U = U^*\Lambda\bar{M}U = U^*\bar{M}\Lambda U\\ &= U^*M\bar{U}U^T\Lambda\bar{U} = (U^*M\bar{U})(\overline{U^*\Lambda U}) = B\bar{A}.\end{aligned}$$

反之，如果 $A\bar{B}=B\bar{A}$，则对某个 $\varepsilon>0$，$A_\varepsilon\equiv A+\varepsilon I$ 是非奇异的 Hermite 矩阵，且 $A_\varepsilon\bar{B}=A\bar{B}+\varepsilon\bar{B}=B\bar{A}+\varepsilon B=B\bar{A}_\varepsilon$. 因此，(4.5.15)的条件Ⅲ(a)(1)被满足，且存在酉矩阵 $U_\varepsilon\in M_n$ 使得 $U_\varepsilon A_\varepsilon U_\varepsilon^*=U_\varepsilon AU_\varepsilon^*+\varepsilon I=\Lambda_\varepsilon$ 和 $U_\varepsilon BU_\varepsilon^T=M_\varepsilon$ 都是对角矩阵，因而 $U_\varepsilon AU_\varepsilon^*=\Lambda_\varepsilon-\varepsilon I$ 和 $U_\varepsilon BU_\varepsilon^T=M_\varepsilon$ 都是对角矩阵. □

两个奇异 Hermite 矩阵经（不一定是酉的）*相合同时对角化的问题在习题 8 中讨论.

我们已经看到，在 *相合下，一个 Hermite 矩阵总可以取非常简单的形式（在对角线上有 ±1 或 0 的对角矩阵），并且，在一定的条件下，一对 Hermite 矩阵经 *相合可以同时变成对角矩阵. 于是，自然要提出的问题是：一般的 Hermite 矩阵偶 A，$B\in M_n$ 在同时 *相合下可以变成什么样的标准形？即经 C 一次相合，矩阵偶

$$C^*AC \text{ 和 } C^*BC$$

可以取什么样的标准形？虽然这个问题是针对（可能都是奇异的）一般 Hermite 矩阵偶来讨论的，但是不论是提出还是证明其一般结果都是相当复杂的. 这里，对其中至少有一个矩阵是非奇异的 Hermite 矩阵偶，不加证明地叙述标准形偶定理. 我们已经讨论了可经 *相合同时对角化的特殊情形.

236

4.5.19 定理 假定 A，$B\in M_n$ 是 Hermite 矩阵，且 A 是非奇异矩阵. 则存在正整数 k 和非奇异矩阵 $C\in M_n$，使得

$$C^*AC=\begin{bmatrix}A_1 & & & 0\\ & A_2 & & \\ 0 & & \ddots & \\ & & & A_k\end{bmatrix},\quad C^*BC=\begin{bmatrix}B_1 & & & 0\\ & B_2 & & \\ 0 & & \ddots & \\ & & & B_k\end{bmatrix},$$

其中，每对 A_i，$B_i\in M_{n_i}$，$i=1, 2, \cdots, k$，是两种可能形式之一：

$$B_i=\epsilon\begin{bmatrix}0 & & & \alpha\\ & & \cdot^{\cdot^{\cdot}} & 1\\ & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \\ \alpha & 1 & & 0\end{bmatrix},\quad A_i=\epsilon\begin{bmatrix}0 & & 1\\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & 0\end{bmatrix}, \tag{4.5.20}$$

且 α 是实数，或

$$B_i=\left[\begin{array}{ccc|ccc} & & & 0 & & \alpha\\ & 0 & & & \cdot^{\cdot^{\cdot}} & 1\\ & & & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \\ & & & \alpha & 1 & 0\\ \hline 0 & & \bar{\alpha} & & & \\ & \cdot^{\cdot^{\cdot}} & 1 & & 0 & \\ \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & & & & \\ \bar{\alpha} & 1 & 0 & & & \end{array}\right],\quad A_i=\begin{bmatrix}0 & & 1\\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & 0\end{bmatrix}, \tag{4.5.21}$$

且 α 是复数. 在(4.5.20)中 ϵ 是 +1 或 −1. 而在(4.5.21)中 n_i 是偶数且两个非零子块都在 $M_{(1/2)n_i}$ 中.

说明：

1. 在 α 是实数的情形，可能有 $n_i=1$，于是两个子块具有形式 $\pm\alpha$，±1. 相应于同一个 α 值(例如还相应于同一个值 $\epsilon=1$)的多个 1×1 子块，在 C^*BC 中将产生形如 αI 的子块，而在 C^*AC 中则是 I.

237

2. 在 α 是复数的情形，可能 $n_i=2$，于是两个子块具有形式

$$B_i=\begin{bmatrix}0 & \alpha\\ \bar{\alpha} & 0\end{bmatrix},\quad A_i=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}.$$

3. 在定理中，同时产生的子块结构恰好对应于 A^{-1} 的 Jordan 标准形. 即 $A^{-1}B$ 的诸基本 Jordan 块恰好是 $A_i^{-1}B_i$. 注意到 $(C^*AC)^{-1}(C^*BC)=C^{-1}(A^{-1}B)C$，因而 C 也是使 $A^{-1}B$ 取 Jordan 标准形的相似矩阵. 于是定理所确认的形式可以从 $A^{-1}B$ 的 Jordan 标准形求得(确定诸惯性因子 ϵ 是不难的).

4.5.22 注释 就像两个 Hermite 矩阵 A，B 在*相合下的标准形偶(4.5.19)类似于 $A^{-1}B$ 的 Jordan 标准形一样，关于两个实对称矩阵 A，B，在实^{T}相合下也存在一个标准形偶，它类似于 $A^{-1}B$ 的实 Jordan 标准形. 其中，形如(4.5.21)的子块 B_i 用形如(3.4.4)的类似子块来代

替，而其他可能形式的子块取原来的形式.

习题

1. 设 A，$B\in M_n$，且假定 B 是非奇异矩阵. 证明存在 $C\in M_n$ 使得 $A=BC$. 此外，对任一非奇异矩阵 $S\in M_n$，有 $SAS^*=(SBS^*)C'$，其中 C' 相似于 C.

2. Sylvester 惯性定律(4.5.8)的证明中较难理解的部分是要证明，若 D_1 和 D_2 是相合的 $n\times n$ 惯性矩阵(4.5.7)，则它们有相同个数的正对角元. 正文中给出的证明依赖于 Courant-Fischer 定理的推论. 请对下述初等证明作详细的论述. 假定 $D_2=S^*D_1S$，并且假定 D_1 恰好有 s 个正对角元且至少有一个负对角元. 假定 D_1 的前 s 个对角元和 D_2 的前 t 个对角元是正的，其中 $1\leqslant s$，$t<n$. 如果 $s<t$，证明存在一个非零向量 $x=[x_i]\in\mathbf{C}^n$ 使得 $x_{t+1}=x_{t+2}=\cdots=x_n=0$ 及 $(Sx)_1=(Sx)_2=\cdots=(Sx)_s=0$. 然后证明 $x^*D_2x>0$ 而 $(Sx)^*D_1(Sx)<0$ 从而得出矛盾.

3. 设 A，$B\in M_n$ 都是 Hermite 矩阵. 证明下列四个条件等价：(a) A 和 B 可经 *相合同时对角化. (b)对某两个非零实纯量 a，b，$aA+bB$ 与 B 可经 *相合同时对角化. (c) A 和 B 同时 *相合于一对可交换的矩阵. (d) $A+iB$ *相合于正规矩阵. [238]

4. 试用定理(4.5.15)证明中的证明方法以及交换族定理(1.3.19)和(4.1.6)证明定理(4.5.15)之Ⅰ(b)的如下推广. 设 A_1，A_2，$\cdots$，$A_k\in M_n$ 是给定的 Hermite 矩阵，其中 A_1 是非奇异的，则存在一个非奇异矩阵 $T\in M_n$ 使得对所有 $i=1$，2，$\cdots$，k，T^*A_iT 是对角矩阵，当且仅当(a)对所有 $i=2$，$\cdots$，k，$A_1^{-1}A_i$ 相似于实对角矩阵，且(b) $\{A_1^{-1}A_i$：$i=2$，$\cdots$，$n\}$ 是一个矩阵交换族. **提示**：设 $C_i=A_1^{-1}A_i$ 且对于所有 $i=2$，$\cdots$，$k\}$，SC_iS^{-1} 是实对角矩阵. 设 $B_i=(S^*)^{-1}A_iS^{-1}$，然后证明 $\{B_i\}$ 一个 Hermite 矩阵交换族. 存在一个酉矩阵 U 使得对所有 $i=2$，$\cdots$，k，UB_iU^* 是对角矩阵，而 $T=US$ 是所要求的相合矩阵. 相应于(4.5.15)之Ⅱ(b)的推广是什么？

5. 由(4.0.4)给出的具有实对称系数矩阵 $A(x)=[a_{ij}(x)]$ 的微分算子 L 在点 $x\in D\subset\mathbf{R}^n$ 是椭圆型的，是指其系数矩阵 $A(x)$ 是非奇异的且它的所有特征值有相同的符号. 称 L 在 x 是双曲型的，是指 $A(x)$ 是非奇异矩阵，且它的 $n-1$ 个特征值有相同的符号，而一个特征值有相反的符号. 试说明，为什么关于一个坐标系一个微分算子在一个点是椭圆型(或双曲型)，则关于其他每个坐标系，这个微分算子在那个点也是椭圆型(或双曲型). Laplace 方程

$$\nabla^2 f=\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}+\frac{\partial^2 f}{\partial z^2}=0$$

给出了椭圆微分算子的一个例子，而波动方程

$$\Box^2 f=\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}-\frac{\partial^2 f}{\partial t^2}=0$$

是双曲型算子的一个例子. 这两个方程都是在笛卡儿坐标系中给出的. 在球极坐标，柱面坐标或其他坐标系下，这两个方程的差别就很大.

6. 设 $X=[X_1，\cdots，X_n]^T$ 和 $Y=[Y_1，\cdots，Y_n]^T$ 是由具有有限二阶矩的实随机变量组成的两个向量. 事实上(见第 7 章)，X 和 Y 的协方差矩阵都只有非负特征值. 假定其中至少一个协方差矩阵是非奇异的. 证明存在实非奇异矩阵 $S\in M_n$，使得 SX 和 SY 的协方差矩阵都是对角矩阵. 用统计学术语表述是，可以求得一个非奇异线性变换 S 使 SX 和 SY 的诸分量各不 [239]

相关.

7. 利用习题 4 给出三个或多个随机向量满足什么条件就能保证有一个非奇异线性变换使变换后的诸随机向量的各分量是不相关的.

8. 定理(4.5.15)的情形 I (b)考虑了两个 Hermite 矩阵在至少有一个矩阵是非奇异的情形下经*相合同时对角化的问题. 推论(4.5.18a)考虑了两个矩阵可能都是奇异的情形下用酉*相合同时对角化的问题. 若两个矩阵都是奇异的，则经(不一定是酉的)*相合同时对角化它们的问题最终可化成(4.5.15)，但必须考察这两个矩阵的两个零空间之交的正交补的性质. 设 A, $B\in M_n$ 是 Hermite 矩阵，且假定它们都是奇异的. 设 $N(A)$和 $N(B)$分别表示 A 和 B 的零空间. (a)考察$\begin{bmatrix}0&0\\0&1\end{bmatrix}$和$\begin{bmatrix}1&0\\0&0\end{bmatrix}$可以证明，存在一对奇异 Hermite 矩阵可经*相合同时对角化. (b)假定 $N(A)\cap N(B)=\{0\}$. 证明，若 A 和 B 可经*相合同时对角化，则存在一个实数 a 使得 $aA+B$ 是非奇异的. **提示**：若 $C\in M_n$ 是非奇异的，$C^*AC=\Lambda_1$，且 $C^*BC=\Lambda_2$，其中 Λ_1 和 Λ_2 是对角矩阵，证明 Λ_1 和 Λ_2 的零主对角元不会处在相同位置. 你能选取 a 使得 $a\Lambda_1+\Lambda_2$ 的主对角元都不为零吗？(c)利用(b)证明

$$A=\begin{bmatrix}0&1&0\\1&0&0\\0&0&0\end{bmatrix}\quad 和\quad B=\begin{bmatrix}0&0&0\\0&0&1\\0&1&0\end{bmatrix}$$

不能经*相合同时对角化. (d)若

$$N(A)\cap N(B)=\{0\},$$

又 $a\in\mathbf{R}$ 不为零且 $aA+B$ 是非奇异的，利用习题 3(b)，A 和 B 可经*相合同时对角化当且仅当 $(aA+B)^{-1}B$ 可对角化且只有实特征值. (e)若 $\dim N(A)\cap N(B)=k\geqslant 1$，设$\{u_1, u_2, \cdots, u_n\}$是 $\mathbf{R}^n$ 的一个标准正交基，而其中的$\{u_1, u_2, \cdots, u_k\}$是 $N(A)\cap N(B)$的一个标准正交基. 若 $U=[u_1, u_2, \cdots, u_n]\in M_n$，证明

240

$$U^*AU=\begin{bmatrix}0&0\\0&A'\end{bmatrix}\quad 和\quad U^*BU=\begin{bmatrix}0&0\\0&B'\end{bmatrix},$$

其中 A', $B'\in M_{n-k}$，$N(A')\cap N(B')=\{0\}$，而左上角的零子块是 $k\times k$ 的. 证明，A 和 B 可经*相合同时对角化当且仅当 A'和 B'可经*相合同时对角化. 虽然 A'和 B'可能都是奇异的，但它们的零空间之交是平凡的. (f)试收集从(a)到(e)的信息来叙述并证明关于两个 Hermite 矩阵经*相合同时对角化的一般定理.

9. 如果 A, $B\in M_n$，且 B 非奇异，证明 A 与 B 可交换，当且仅当 A 与 B^{-1}可交换.

10. 证明$\begin{bmatrix}0&1\\1&0\end{bmatrix}$和$\begin{bmatrix}1&0\\0&-1\end{bmatrix}$可经酉^{T}相合同时化简成对角形式，但不能经*相合同时化简成对角形式. 试用(4.5.15)情形 II (b)证明中所采用的构造法实现化简，且顺便求出实施这个化简的酉^{T}相合矩阵.

11. 证明$\begin{bmatrix}1&1\\1&0\end{bmatrix}$和$\begin{bmatrix}0&1\\1&0\end{bmatrix}$不能经*相合或^{T}相合同时化简成对角形式.

12. 设 A, $B\in M_n$, 且 A 非奇异. 证明, 下列条件中的每一个, 是 A 和 B 在定理(4.5.15)(表 4.5.15T)所指每种情形的假设下经相应意义下的相合同时对角化的必要充分条件.

情形	必要充分条件
Ⅰ(a)	存在 Hermite 矩阵 $F\in M_n$ 使得 $B=AF$
Ⅰ(b)	存在具有实特征值的可对角化矩阵 $F\in M_n$ 使得 $B=AF$
Ⅱ(a)	存在正规矩阵 $F\in M_n$ 使得 $B=AF$
Ⅱ(b)	存在可对角化矩阵 $F\in M_n$ 使得 $B=AF$
Ⅲ(a)	存在对称矩阵 $F\in M_n$ 使得 $B=AF$
Ⅲ(b)	存在可合对角化矩阵 $F\in M_n$ 使得 $B=AF$[见(4.6.2)]

13. 设 A, $B\in M_n$ 是对称矩阵(可能都是奇异矩阵), 且假定存在酉矩阵 $U\in M_n$ 使得 $UAU^T=\Lambda$ 和 $UBU^T=M$ 都是对角矩阵. 证明存在酉矩阵 V 使得 $B\bar{A}=AVB$. **提示:** 如果 $\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, 证明存在酉对角矩阵 D 使得 $\bar{\Lambda}=D\Lambda=\Lambda D$. 然后证明 241

$$B\bar{A}=U^T M\bar{\Lambda}\bar{U}=U^T\Lambda D_1D_2\bar{M}\bar{U}=A(U^TD_1D_2\bar{U})\bar{B},$$

其中 D_1 和 D_2 是酉对角矩阵.

14. 利用习题 13 的必要条件证明, 习题 8(c)的两个对称矩阵不能经酉^{T}相合同时对角化. **提示:** 计算 $B\bar{A}$ 和 AUB 的第一列. 利用(4.5.18b)证明同样的结论更容易.

15. 如果 A, $B\in M_n$ 是对称矩阵, 证明, 只要 A 和 B 都是非奇异矩阵, 习题 13 中经^{T}相合同时对角化的必要条件也是充分条件. **提示:** 如果 $B\bar{A}=AUB$, 且 A 和 B 是非奇异矩阵, 则 $A^{-1}B\bar{A}\bar{B}^{-1}=U$ 且 $I=UU^*$. 这便推出 $\bar{A}\bar{B}^{-1}B^{-1}A=B^{-1}A\bar{A}\bar{B}^{-1}$. 两边取逆推出 $A^{-1}B$ 是正规矩阵.

16. 设 A, $B\in M_n$ 是对称矩阵(可能都是奇异矩阵), 并且假定存在一个酉矩阵 $U\in M_n$ 使得 $UAU^T=\Lambda$ 和 $UBU^T=M$ 都是对角矩阵. 证明 $A\bar{A}$ 与 $B\bar{B}$ 可交换. 通过考察习题 8(c)中的两个矩阵说明, 经酉^{T}相合同时对角化的上述必要条件不是充分条件. 用推论(4.4.5)证明, 这个必要条件是充分条件, 只要 $A\bar{A}$ 和 $B\bar{B}$ 都有 n 个不同的特征值.

17. 设 A, $B\in M_n$, A 是 Hermite 矩阵, B 是对称矩阵, 又假定存在酉矩阵 $U\in M_n$ 使得 $UAU^*=\Lambda$ 和 $UBU^T=M$ 都是对角矩阵. 证明 A 与 $B\bar{B}$ 可交换. 通过考察习题 11 中的两个矩阵说明这个可经(* 和^T)相合同时对角化的必要条件不是充分条件. 试用推论(4.4.5)证明, 只要 $B\bar{B}$ 的所有特征值互不相同, 则这个必要条件也是充分条件.

18. 设 A, $B\in M_n$ 且 A 和 B 是对称矩阵, A 还是非奇异矩阵. 证明, 如果广义特征多项式 $p_{A,B}(t)\equiv\det(tA-B)$ 有 n 个不同的零点, 则 A 和 B 可经^{T}相合同时对角化. **提示:** $A^{-1}B$ 的诸特征值是什么?

19. 对 Sylvester 惯性定律(4.5.8)的下述另一个证明作详细的论述. 若 $A\in M_n$ 是一个非奇异 Hermite 矩阵, 且 $S\in M_n$ 是非奇异矩阵. 设 $S=QR$ 是一个分解, 其中 $Q\in M_n$ 是酉矩阵, 而 $R\in M_n$ 是一个具有正主对角元的上三角矩阵(见 2.6.1). 证明, 若 $0\leqslant t\leqslant 1$, 则 $S(t)=tQ+(1-t)QR$ 是非奇异矩阵. 设 $A(t)\equiv S(t)AS(t)^*$, $A(0)$和 $A(1)$是什么矩阵? 因为 $A(t)$是非奇异矩阵, 且当 t 由 0 变到 1 时, $A(t)$是连续地变化. 证明 $A(0)$与 $A(1)$有相同个数的正(负)特 242

征值. 对较小的 $\varepsilon>0$ 考察 $A\pm\varepsilon I$，并论述一般的情形.

20. 如果 $A=\begin{bmatrix} B & C \\ 0 & 0 \end{bmatrix}\in M_n$，且 $B\in M_k$，$1\leqslant k\leqslant n$. 证明 A 是正规矩阵，当且仅当 B 是正规矩阵且 $C=0$. **提示**：计算 AA^* 和 A^*A，如果 $C^*C=0$，则对于所有 $x\in\mathbf{C}^{n-k}$，$(Cx)^*(Cx)=0$，因而对所有 $x\in\mathbf{C}^{n-k}$，$Cx=0$.

21. 说明(4.5.18)(b)中所采用的证法也可以用来证明(a)和(c)这两部分.

22. 设 $\mathscr{F}=\{A_1,\cdots,A_k\}\subset M_n$ 是给定的复对称矩阵族，又设 $\mathscr{G}=\{A_i\bar{A}_j: i, j=1, 2, \cdots, k\}$. 如果存在一个酉矩阵 $U\in M_n$ 使得对所有的 $i=1,\cdots,k$，UA_iU^T 是对角矩阵，证明 $\mathscr{G}$ 是一个交换族. 当 $k=2$ 时这简化成什么结论，且与(4.5.18b)有什么关系？事实上，$\mathscr{G}$ 的交换性也足以确保 $\mathscr{F}$ 经酉^{T}相合可同时对角化；请参看本节末的"进一步阅读"中所引用的 Hong 和 Horn 的文章.

23. 设 $\mathscr{F}=\{A_1,\cdots,A_k\}\subset M_n$ 是给定的复对称矩阵族，$\mathscr{H}=\{B_1,\cdots,B_m\}\subset M_n$ 是给定的 Hermite 矩阵族，又设 $\mathscr{G}=\{A_i\bar{A}_j: i, j=1,\cdots,k\}$. 如果存在一个酉矩阵 $U\in M_n$ 使得每个 UA_iU^T 和每个 UB_jU^* 是对角矩阵. 证明 $\mathscr{G}$ 和 $\mathscr{H}$ 都是交换族. 且对于所有 $i=1,\cdots,k$ 和所有 $j=1,\cdots,m$，B_jA_i 是对称矩阵. 当 $k=m=1$ 时，这简化成什么结论，且与(4.5.18c)有什么关系？事实上，这些条件也足以确保 $\mathscr{F}$ 和 $\mathscr{H}$ 分别经相合同时对角化. 请参看本节末的"进一步阅读"中所引用的 Hong 和 Horn 的文章.

进一步阅读 定理(4.5.9)的 Ostrowski 的证明以及有关的结果可参看"A Quantitative Formulation of Sylvester's Law of Inertia," *Proc. Nat. Acad. Sci.* 45(1959), 740-744. 定理(4.5.25)的另一种形式在[GLR 82]中给出；包括两个矩阵是奇异矩阵的情形的一个详细证明还在 R. C. Thompson 未发表的手稿中. 有关两个以上矩阵同时对角化的结果可参看 Y. P. Hong and R. A. Horn, "On Simultaneous Reduction of Families of Matrices to Triangular or Diagonal Form by Unitary Congruence," *Linear and Multilinear Algebra* 17(1985), 271-288.

243

4.6 合相似和合对角化

提出本节论题的动机来源于前两节的三个结果. 定理(4.4.3)刻化了所有形如 $U\Delta U^T$ 的矩阵，其中，Δ 是上三角矩阵而 U 是酉矩阵；为了现在的目的，需要把这个分解写成 $U\Delta U^T=U\Delta\bar{U}^{-1}$. 推论(4.4.4)刻划了所有形如 $U\Sigma U^T=U\Sigma\bar{U}^{-1}$ 的矩阵，其中 Σ 是对角矩阵，而定理(4.5.15)的情形Ⅲ则要求下述知识：何时一个给定的复方阵 A 可经变换 $A\to SA\bar{S}^{-1}$ 化简成对角形式，其中 S 为某个非奇异方阵.

4.6.1 定义 设矩阵 A，$B\in M_n$，如果存在非奇异矩阵 $S\in M_n$ 使得 $A=SB\bar{S}^{-1}$，就称 A 和 B 合相似(Consimilar). 如果矩阵 S 可以取酉矩阵，则称 A 和 B 酉合相似.

如果 $A=SB\bar{S}^{-1}$，且 $S=U$ 是酉矩阵，则 $A=SB\bar{S}^{-1}=UBU^T$；如果 $S=Q$ 是复正交矩阵，则 $A=SBS^{-1}=QBQ^*$；如果 $S=R$ 是非奇异实矩阵，则 $A=SB\bar{S}^{-1}=RBR^{-1}$. 因此，合相似的各种特殊情形包括^{T}相合，*相合和普通的相似.

像普通相似一样，合相似是 M_n 上的等价关系，因而我们可能要问，哪些等价类包含三角代表元或对角代表元.

4.6.2 定义 设矩阵 $A\in M_n$，如果存在非奇异矩阵 $S\in M_n$ 使得 $S^{-1}A\bar{S}$ 是上三角矩阵，则称 A 可合三角化；如果可选取 S 使得 $S^{-1}A\bar{S}$ 是对角矩阵，就称 A 可合对角化；称 A 可酉合三角化或可酉合对角化，是指 A 可以通过酉矩阵的合相似化简成所要求的形式.

如果 $A\in M_n$ 可合三角化，并且 $S^{-1}A\bar{S}=\Delta$ 是上三角矩阵，则由直接计算可知，$\Delta\bar{\Delta}=S^{-1}(A\bar{A})S$ 的诸主对角元都是非负的. 因而，$A\bar{A}$ 的所有特征值都是非负的. 另一方面，定理(4.4.3)说明，存在酉矩阵 U 使得 $UAU^T=UA\bar{U}^{-1}$ 是上三角矩阵. 因此，确定某个矩阵是否可经合相似化简成上三角矩阵的问题已经解决了.

4.6.3 定理 设 $A\in M_n$ 是给定的矩阵. 则下列命题等价： [241]

(a) A 可合三角化；

(b) A 可酉合三角化；

(c) $A\bar{A}$ 的所有特征值都是非负实数.

如果 $A\in M_n$ 是可酉合对角化的，则对某个酉矩阵 $U\in M_n$ 和 $\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_n)$，$A=U\Lambda\bar{U}^{-1}=U\Lambda U^T$. 于是 $A^T=(U\Lambda U^T)^T=U\Lambda^TU^T=U\Lambda U^T=A$，因而 A 是对称矩阵. 推论(4.4.4)说明，逆命题也成立，且对角矩阵总可以取非负的. 因此，我们又解决了酉合对角化的问题.

4.6.4 定理 矩阵 $A\in M_n$ 可酉合对角化，当且仅当 A 是对称矩阵.

其余与合三角化和合对角化有关的问题是有效地刻划那些可经一个不一定是酉的合相似的可合对角化矩阵.

如果 $A\in M_n$ 可合对角化且 $S^{-1}A\bar{S}=\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_n)$，则 $A\bar{S}=S\Lambda$. 如果 $S=[s_1\cdots s_n]$，其中每个 $s_i\in\mathbf{C}^n$，这个恒等式说明 $A\bar{s}_i=\lambda_i s_i$ 对 $i=1,\cdots,n$ 成立. 这个方程类似于通常的特征向量-特征值方程，但又与它有本质的差别.

4.6.5 定义 设 $A\in M_n$ 是给定的矩阵. 如果对某个 $\lambda\in\mathbf{C}$，非零向量 $x\in\mathbf{C}^n$ 适合 $A\bar{x}=\lambda x$，就称 x 为 A 的合特征向量；纯量 λ 是 A 的合特征值.

恒等式 $A\bar{S}=S\Lambda$ 表明，S 的每个非零列是 A 的合特征向量. 因为 S 的诸列无关当且仅当 S 非奇异，所以得知，矩阵 $A\in M_n$ 可合对角化，当且仅当它有 n 个无关的合特征向量. 从这个意义上讲，合对角化理论完全类似于普通的对角化理论.

但是，每个矩阵至少有一个特征值，且它只有有限多个不同的特征值；从这方面考虑，合特征值理论则大不相同. 如果 $A\bar{x}=\lambda x$，则 $e^{-i\theta}A\bar{x}=A(\overline{e^{i\theta}x})=e^{-i\theta}\lambda x=(e^{-2i\theta}\lambda)(e^{i\theta}x)$ 对所有 $\theta\in\mathbf{R}$ 成立. 因此，如果 λ 是 A 的合特征值，则对所有 $\theta\in\mathbf{R}$，$e^{i\theta}\lambda$ 亦是 A 的合特征值. 另一方面，如果 $A\bar{x}=\lambda x$，则 $A\bar{A}x=A(\overline{A\bar{x}})=A(\overline{\lambda x})=\bar{\lambda}A\bar{x}=\bar{\lambda}\lambda x=|\lambda|^2x$，因而，只有当 $|\lambda|^2$ 是 $A\bar{A}$ 的特征值时，这个纯量 λ 才可能是 A 的合特征值. 例如 $A=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$ 适合 $A\bar{A}=-2I$，而 $A\bar{A}$ 没有非负特征值，这个例子说明，有些矩阵根本没有合特征值. 但是，已经知道，如果 $A\in M_n$，

245 且 n 是奇数，则 A 至少必有一个合特征值，这个结果类似于每个奇数阶实矩阵至少有一个实特征值的事实.

因此，与普通的特征值理论相反，一个矩阵可以有无限多个不同的合特征值，或者它可能根本没有合特征值. 如果一个矩阵有合特征值，有时为了方便，从模相同的合特征值中选出唯一的非负特征值作为代表.

刚才得到的关于合特征值存在的必要条件也是充分条件.

4.6.6 命题 设 $A\in M_n$，且 $\lambda\geqslant 0$ 是给定的，则 λ 是 $A\bar{A}$ 的特征值，当且仅当 $+\sqrt{\lambda}$ 是 A 的合特征值.

证明：如果 $\lambda\geqslant 0$，$\sqrt{\lambda}\geqslant 0$，且对某个 $x\neq 0$ 有 $A\bar{x}=\sqrt{\lambda}x$，则 $A\bar{A}x=A(\overline{A\bar{x}})=A(\overline{\sqrt{\lambda}x})=\sqrt{\lambda}A\bar{x}=\sqrt{\lambda}\sqrt{\lambda}x=\lambda x$.

反过来，如果对某个 $x\neq 0$，$A\bar{A}x=\lambda x$，则有两种可能情形：

(a) $A\bar{x}$ 和 x 相关；

(b) $A\bar{x}$ 和 x 无关.

在前一种情形，存在某个 $\mu\in\mathbf{C}$ 使得 $A\bar{x}=\mu x$，这说明 μ 是 A 的合特征值. 另一方面，$\lambda x=A\bar{A}x=A(\overline{A\bar{x}})=A(\overline{\mu x})=\bar{\mu}A\bar{x}=\bar{\mu}\mu x=|\mu|^2x$，所以 $|\mu|=+\sqrt{\lambda}$. 因为对任意 $\theta\in\mathbf{R}$，$e^{-2i\theta}\mu$ 是相应于合特征向量 $e^{i\theta}x$ 的合特征值，由此得出 $+\sqrt{\lambda}$ 是 A 的合特征值. 注意到 $A\bar{A}(A\bar{x})=A(\overline{A\bar{A}x})=A(\overline{\lambda x})=\lambda(A\bar{x})$ 及 $A\bar{A}x=\lambda x$，因而，如果 λ 是 $A\bar{A}$ 的单特征值，情形(a)一定总会出现.

在后一种情形(b)(如果 λ 是 $A\bar{A}$ 的重特征值，就可能出现这种情形)，向量 $y=A\bar{x}+\sqrt{\lambda}x$ 是非零的，又因为

$$A\bar{y}=A\bar{A}x+\sqrt{\lambda}A\bar{x}=\lambda x+\sqrt{\lambda}A\bar{x}=\sqrt{\lambda}(A\bar{x}+\sqrt{\lambda}x)=\sqrt{\lambda}y,$$

所以 y 是相应于合特征值 $+\sqrt{\lambda}$ 的合特征向量. □

我们已经看到，对 $A\bar{A}$ 的每个不同的非负特征值，都有 A 的一个相应的合特征向量，这个结果类似于普通的特征向量理论. 下述结果又稍微推广了这种类似性.

4.6.7 命题 设 $A\in M_n$ 是给定的矩阵，又设 $x_1, x_2, \cdots, x_k$ 是 A 的相应于合特征值 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 的合特征向量. 如果当 $1\leqslant i, j\leqslant k$ 且 $i\neq j$ 时，$|\lambda_i|\neq|\lambda_j|$，则 $\{x_1, x_2, \cdots, x_k\}$ 是
246 线性无关向量组.

证明：每个 x_i 是 $A\bar{A}$ 的相应于特征值 $|\lambda_i|^2$ 的特征向量. 由于向量 $x_1, x_2, \cdots, x_k$ 是矩阵 $A\bar{A}$ 的特征向量，且根据假设它们的特征值 $|\lambda_1|^2, \cdots, |\lambda_k|^2$ 是两两不同的，所以由(1.3.8)可知它们是无关的. □

这个结果连同命题(4.6.6)给出了一个已知矩阵的无关合特征向量个数的下界，由此得出可合对角化的充分条件，这类似于我们所熟悉的普通的可对角化的充分条件. 在定理(4.6.11)中要给出更一般的条件.

4.6.8 推论 设 $A\in M_n$ 是给定的矩阵. 如果 $A\bar{A}$ 有 k 个互异的非负特征值，则 A 至少有 k 个无关的合特征向量. 如果 $k=n$，则 A 可合对角化. 如果 $k=0$，则 A 根本没有合特征向量.

关于无关的合特征向量的个数的这些界是可以达到的. 对于 $A=J_n(1)$，这是一个基本

Jordan 块

$$J_n = \begin{bmatrix} 1 & 1 & & 0 \\ & 1 & \ddots & \\ & & \ddots & 1 \\ 0 & & & 1 \end{bmatrix} \in M_n,$$

$A\bar{A}=J_n^2(1)$ 以 1 作为它唯一的非负特征值. 易知合特征向量方程 $A\bar{x}=x$ 只有实解，因而每个合特征向量也是特征向量，且特征向量组成的子空间是一维的. 因此，对任意适合 $1\leqslant k\leqslant n$ 的整数 k，可以用基本 Jordan 块的直和给出这样一个矩阵 $A\in M_n$ 的例子，使得 $A\bar{A}$ 有 k 个不同的非负特征值且 A 恰好有 k 个无关合特征向量.

我们的目的是要给出一个使给定的矩阵可合对角化的简单条件，作为第一步，先证明下述引理. 提出这个结果是因为，如果某个矩阵 $A\in M_n$ 合相似于一个纯量矩阵，则 $A=S(\lambda I)\bar{S}^{-1}=\lambda S\bar{S}^{-1}$ 且 $A\bar{A}=\lambda S\bar{S}^{-1}\bar{\lambda}\bar{S}S^{-1}=|\lambda|^2 I$. 具有($A\bar{A}$ 是纯量矩阵)这个性质的矩阵是构成可合对角化矩阵的基本子块.

4.6.9 引理 矩阵 $A\in M_n$ 有性质 $A\bar{A}=I$，当且仅当存在非奇异矩阵 $S\in M_n$ 使得 $A=S\bar{S}^{-1}$.

证明：我们刚才已经看到，所述条件是必要的. 为了证明它是充分的，对任意 $\theta\in\mathbf{R}$，定义 $S_\theta=e^{i\theta}A+e^{-i\theta}I$，注意到

$$A\bar{S}_\theta = A(e^{-i\theta}\bar{A}+e^{i\theta}I) = e^{-i\theta}A\bar{A}+e^{i\theta}A = e^{i\theta}A+e^{-i\theta}I = S_\theta. \tag{4.6.10}$$

因为 A 只有有限多个特征值，所以存在某个 $\theta\in\mathbf{R}$ 使得 $-e^{-2i\theta_0}$ 不是 A 的特征值. 对于 θ 的这个值， [247]

$$S_{\theta_0} = e^{i\theta}(A+e^{-2i\theta_0}I)$$

是非奇异矩阵，且由(4.6.10)有 $A=S_{\theta_0}\bar{S}_{\theta_0}^{-1}$. □

我们现在可以叙述并证明可合对角化的必要充分条件了.

4.6.11 定理 设 $A\in M_n$. 则存在非奇异矩阵 $S\in M_n$ 和对角矩阵 $\Lambda\in M_n$ 使得 $A=S\Lambda\bar{S}^{-1}$，当且仅当 $A\bar{A}$ 是具有非负实特征值的可对角化矩阵，且 rank A=rank $A\bar{A}$.

证明：所述条件显然是必要的，因为

$$A\bar{A} = S\Lambda\bar{S}^{-1}\bar{S}\bar{\Lambda}S^{-1} = S|\Lambda|^2S^{-1},$$

且 $A\bar{A}$ 的秩与 A 的秩都是 Λ 中非零对角元的个数. 反之，如果 $A\bar{A}$ 可对角化且有非负特征值，则存在非奇异矩阵 $S\in M_n$ 和非负对角矩阵 $\Lambda\in M_n$ 使得 $A\bar{A}=S\Lambda S^{-1}$. 不失一般性，假定 Λ 中的相同对角元都排放在一起，且 $\Lambda=\lambda_1 I_{n_1}\oplus\lambda_2 I_{n_2}\oplus\cdots\oplus\lambda_k I_{n_k}$，其中 $I_{n_i}\in M_{n_i}$ 且 $\lambda_1>\lambda_2>\lambda_3>\cdots>\lambda_k\geqslant 0$. 于是有

$$S^{-1}A\bar{A}S = S^{-1}A\bar{S}\bar{S}^{-1}\bar{A}S = (S^{-1}A\bar{S})(\overline{S^{-1}A\bar{S}}) = \Lambda.$$

如果令 $B=S^{-1}A\bar{S}$，则(因为合相似是等价关系)只需证明：若 $B\bar{B}=\Lambda$，B 就可对角化. 因为 Λ 是实矩阵，$\Lambda=\bar{\Lambda}=(\overline{B\bar{B}})=\bar{B}B=B\bar{B}$，所以 B 与 $\bar{B}$ 可交换. 于是 $B\Lambda=B(\bar{B}B)=B\bar{B}B=(B\bar{B})B=\Lambda B$. 所以 B 与 Λ 也可交换. 如果把 B 表示成分块形式

$$B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1k} \\ \vdots & B_{22} & & \vdots \\ & & \ddots & \\ B_{k1} & \cdots & & B_{kk} \end{bmatrix},$$

其中子块的阶数与

$$A=\begin{bmatrix}\lambda_1 I_{n_1} & & 0\\ & \ddots & \\ 0 & & \lambda_k I_{n_k}\end{bmatrix},\ I_{n_i}\in M_{n_i},\ i=1,2,\cdots,k$$

[218]

的相应子块相同，则方程 $BA=AB$ 表明，对所有 $i=1,2,\cdots,k$ 有 $\lambda_i B_{ij}=\lambda_j B_{ij}$. 因为如果 $i\neq j$，则 $\lambda_i\neq\lambda_j$ 推出，如果 $i\neq j$，则 $B_{ij}=0$，因而 B 是分块对角矩阵

$$B=\begin{bmatrix}B_{11} & & 0\\ & \ddots & \\ 0 & & B_{kk}\end{bmatrix},$$

其中，对角子块与 A 的相应子块有相同的阶数. 方程 $B\bar{B}=A$ 表明，对每个 $i=1,2,\cdots,k$，$B_{ii}\bar{B}_{ii}=\lambda_i$. 注意，如果 $\lambda_i>0$，B_{ii} 必定是非负矩阵. 所以，如果 $\lambda_i>0$，可以把这个方程写成

$$\left[\frac{1}{\sqrt{\lambda_i}}B_{ii}\right]\left[\overline{\frac{1}{\sqrt{\lambda_i}}B_{ii}}\right]=I_{n_i},$$

因而可以利用引理(4.6.9)得出，存在非奇异矩阵 $S_i\in M_{n_i}$，使得 $B_{ii}=S_i(\sqrt{\lambda_i}I_{n_i})\bar{S}_i^{-1}$. 如果 $\lambda_k=0$，则

$$\begin{aligned}&\operatorname{rank} B_{11}+\operatorname{rank} B_{22}+\cdots+\operatorname{rank} B_{kk}\\ &\quad=\operatorname{rank} B=\operatorname{rank} A=\operatorname{rank} A\bar{A}=\operatorname{rank} A=n_1+n_2+\cdots+n_{k-1}.\end{aligned}$$

这表明 $B_{k,k}$ 的秩是零，所以如果 $\lambda_k=0$，则最后一个子块 B_{kk} 实际上必须是零子块. 这时，可以把 B_{kk} 写成 $0=B_{kk}=S_k(\sqrt{\lambda_k}I)\bar{S}_k^{-1}$，其中 $S_k\in M_{n_k}$ 是任意非奇矩阵. 如果令 $S=S_1\oplus\cdots\oplus S_k$，那么所有情形都证明了

$$B=S(\sqrt{\lambda_1}I_{n_1}\oplus\cdots\oplus\sqrt{\lambda_k}I_{n_k})\bar{S}^{-1},$$

这正是想要做的. □

当把可合对角化的必要充分条件应用于定理(4.5.15)的情形Ⅲ(b)时，有下述推论. 如果给定 $A, B\in M_n$，A 是 Hermite 矩阵，B 是对称矩阵，且 A, B 中至少有一个是非奇异矩阵. 按 A 或 B 是非奇异矩阵，令 $C=A^{-1}B$ 或 $B^{-1}A$. 于是，存在非奇异矩阵 $S\in M_n$ 使得 SAS^* 和 SBS^T 都是对角矩阵，当且仅当 $C\bar{C}$ 可角化，其全部特征值是非负的，且 $\operatorname{rank} C=\operatorname{rank} C\bar{C}$.

A 是复对称矩阵的特殊情形容易通过定理(4.6.11)来处理，因为这时 $A\bar{A}=AA^*$ 是 Hermite 矩阵，因而可对角化. 另外，对任意 $A\in M_n$，$\operatorname{rank} A=\operatorname{rank} AA^*$，所以，当 A 是复对称矩阵时，它满足定理的假设. 定理说明每个复对称矩阵可合对角化，但是没有直接得出这时合对角化可经酉变换来实现这个事实. 参看本节末习题 22.

[249]

关于合相似和合对角化的这些论断有助于深入理解关于复对称矩阵的 Takagi 分解(4.4.4)和关于经酉相合三角化的定理(4.4.3). 定理(4.4.3)说明每个使 $A\bar{A}$ 有全部非负特征值的矩阵 $A\in M_n$ 可酉合三角化，而 Takagi 的结果说明每个复对称矩阵可以酉合对角化.

因为对于合特征值，区别"实"和"非实"没有什么用处. 所以在类似于 Hermite(或正定)矩阵的"具有实(或正)合特征值的可酉合对角化"与类似于正规矩阵的"具有复的合特征值的可酉

合对角化"之间就没有什么差别. 因此，复对称矩阵可以看作与整个正规矩阵类(关于普通的相似)类似的矩阵类(关于合相似)，而 Takagi 分解可以看成与正规矩阵的谱定理(2.5.4a，b)类似的结果.

普通的相似性理论的产生是由于研究不同基下的线性变换的结果. 一般说来，合相似的产生是由于研究不同基下的反线性变换的结果. 反线性变换 T 是从一个复向量空间到另一个复向量空间的映射 $T: V\to W$，具有可加性[$T(x+y)=Tx+Ty$ 对所有 $x, y\in V$ 成立]，不过只具有共轭齐次性[$T(ax)=\bar{a}Tx$ 对所有 $a\in\mathbf{C}$ 和所有 $x\in V$ 成立，有时称之为反齐次性]. 在量子力学中，研究时间反转时要出现这样的变换.

合对角化矩阵类是一个广泛的矩阵类. 它包括具有实特征值的所有实可对角化矩阵，所有(实或复)对称矩阵，以及所有形如 H^2S 的矩阵，其中，H 是 Hermite 矩阵，而 S 是对称矩阵(见本节末习题 8 和 9). 后一个论断为下述有用的充分条件中的第二个奠定了基础. 正定矩阵 $A\in M_n$ 是指对所有非零 $x\in\mathbf{C}^n$ 有 $x^*Ax>0$ 的非奇异 Hermite 矩阵；关于一个 Hermite 矩阵 A 是正定的等价条件是，A 的所有特征值都是正，或对于某个非奇异 Hermite 矩阵 H，有 $A=H^2$(见第 7 章).

4.6.12 推论 设 $A, B\in M_n$，且 A 是正定的 Hermite 矩阵.

(a) 如果 B 是 Hermite 矩阵，则存在非奇异矩阵 $S\in M_n$，使得 $SAS^*=I$，且 SBS^* 是实对角矩阵.

(b) 如果 B 是对称矩阵，则存在非奇异矩阵 $S\in M_n$，使得 $SAS^*=I$，且 SBS^T 是具有非负主对角元的实对角矩阵.

250

证明： 设 $A=H^2$，其中 $H\in M_n$ 是非奇异 Hermite 矩阵.

(a) $C\equiv A^{-1}B=H^{-2}B$，所以 C 相似于 $HCH^{-1}=H(H^{-2}B)H^{-1}=H^{-1}BH^{-1}$，这是 Hermite 矩阵，因而具有实特征值且可对角化；矩阵 C 也一定可对角化，且具有实特征值. 因此 A 和 B 可以通过(4.5.15)Ⅰ(b)(2)的相合同时对角化. 如果 $H^{-1}BH^{-1}=U\Lambda U^*$ 其中 U 是酉矩阵且 Λ 是对角矩阵，则非奇异矩阵 $S=U^*H^{-1}$ 将使 $SAS^*=I$ 且 $SBS^*=\Lambda$.

(b) $C\equiv A^{-1}B=H^{-2}B$，所以 $C\bar{C}=H^{-2}B\bar{H}^{-2}\bar{B}$ 相似于

$$H(C\bar{C})H^{-1}=H^{-1}B\bar{H}^{-2}\bar{B}H^{-1}=(H^{-1}B\bar{H}^{-1})(H^{-1}B\bar{H}^{-1})^*,$$

它是 Hermite 矩阵，又是半正定矩阵. 因而它可对角化且具有非负特征值. 根据(0.4.6d)，

$$\operatorname{rank}(C\bar{C})=\operatorname{rank}(H^{-1}B\bar{H}^{-1})(H^{-1}B\bar{H}^{-1})^*=\operatorname{rank}(H^{-1}B\bar{H}^{-1}),$$

再根据(0.4.6b)，$\operatorname{rank}(H^{-1}B\bar{H}^{-1})=\operatorname{rank}(H^{-2}B)=\operatorname{rank} C$. 因此，由(4.6.11)知，(4.5.15)的条件Ⅲ(b)(1)被满足，因而一定有非奇异矩阵 $S\in M_n$，使得 SAS^* 和 SBS^T 都是对角矩阵. 注意到 $HC(H^{-1})^T=H(H^{-2}B)(H^{-1})^T=H^{-1}B(H^{-1})^T$ 是对称矩阵，因此根据(4.4.4)，存在酉矩阵 U 和非奇异对角矩阵 Σ，使得 $H^{-1}B(H^{-1})^T=U\Sigma U^T$ 或 $(U^*H^{-1})B(U^*H^{-1})^T=\Sigma$. 如果令 $S=U^*H^{-1}$，则还有 $S^*AS=I$. □

我们已经讨论了各相似于一个对角矩阵的问题，但是不是每个矩阵都可合对角化. 因而自然要问，在合相似下任一矩阵是否可以化简成某种简单的形式. 在合相似下，有一个标准形，它起的作用类似于 Jordan 标准形在普通相似性中的作用. 利用它，可以证明，对每个 $A\in M_n$，

A 合相似于 $\bar{A}$，A^* 和 A^T[与(3.2.3)比较]，A 合相似于 Hermite 矩阵[与(4.4.9)比较]，A 合相似于实矩阵，并且存在非奇异对称矩阵 S_1，$S_2\in M_n$ 和 Hermite 矩阵 H_1，$H_2\in M_n$ 使得 $A=S_1H_1=H_2S_2$[与推论(4.4.11)比较]. 实际上，可以把整个合相似性的问题归并为一些更熟悉的概念：两个矩阵 A，$B\in M_n$ 合相似的必要充分条件是(a)$A\bar{A}$ 相似于 $B\bar{B}$，且(b)rank $A=$ rank B，rank $A\bar{A}=$ rank $B\bar{B}$，rank $A\bar{A}A=$ rank $B\bar{B}B$，…等对所有 n 个这样的交错乘积均成立，其中乘积的项数最多是 n 个.

习题

1. 证明合相似性是 M_n 上的等价关系.

[251] 2. 给出定理(4.6.3)的证明细节.

3. 设 $A\in M_n$ 是一个给定的矩阵，且设 λ 是 A 的合特征值. 证明 A 的相应于 λ 的合特征向量的集合不一定是 **C** 上 $\mathbf{C}^n$ 的子空间，但它总是 **R** 上的子空间. 试与 A 的普通特征向量的情形相比较.

4. 定理(4.6.11)给出了一个矩阵可合对角化的必要充分条件，但是，当我们考虑多个矩阵时，它们可同时合对角化的条件是什么？设$\{A_1, A_2, \cdots, A_k\}\subset M_n$ 是给定的，并且假定存在一个非奇异矩阵 $S\in M_n$ 使得对 $i=1,\cdots,k$ 有 $\Lambda_i=SA_i\bar{S}^{-1}$，且每个 Λ_i 是对角矩阵. 证明，(a)每个 A_i 可合对角化；(b)每个 $A_i\bar{A}_j$ 可对角化；(c)乘积族$\{A_i\bar{A}_j: i, j=1,\cdots,k\}$是可交换的；(d)对于所有 i，$j=1,\cdots,k$，$A_i\bar{A}_j+A_j\bar{A}_i$ 只有实特征值而 $A_i\bar{A}_j-A_j\bar{A}_i$ 只有虚特征值. 当 $k=1$ 时，这指的是什么？事实上，这些必要条件也是充分条件；关于其证明可参看(4.5)节末的"进一步阅读"中引用的 Hong 和 Horn 的文章.

5. 矩阵 $A\bar{A}$ 在合相似性理论起着重要的作用. 证明，对任一 $A\in M_n$，$A\bar{A}$ 的特征多项式的系数都是实的，由此推出，$A\bar{A}$ 的任何复特征值必须成共轭对出现. **提示：** $\det(tA-A\bar{A}A)=\det A\det(tI-\bar{A}A)=\det(tI-A\bar{A})\det A$. 因此，如果 A 是非奇异矩阵，则 $\bar{A}A$ 和 $A\bar{A}=\overline{(\bar{A}A)}$的特征多项式相同. 对一般情形则考察 $A_\varepsilon=A+\varepsilon I$. 关于 $A\bar{A}$ 的更为明确的结果见习题 8.

6. $A\bar{A}$ 的非负特征值可导出 A 的合特征值. 但是，$A\bar{A}$ 的不是非负的任一特征值也有意义. 假定 $A\in M_n$，且 $A\bar{A}x=\lambda x$ 对某个 $x\neq 0$ 和某个适合 $\lambda\notin[0,\infty)$的 $\lambda\in\mathbf{C}$ 成立. 设 $\alpha\in\mathbf{C}$ 是 λ 的任一平方根，且用 $A\bar{x}=\alpha y$ 定义向量 y. 证明，$A\bar{y}=\bar{\alpha}x$，$A\bar{A}y=\bar{\lambda}y$，及 x 与 y 无关. **提示：** 如果它们相关，x 必须是合特征向量且 $\lambda\geq 0$. 证明 $A\bar{A}$ 的所有复特征值必须成共轭对出现，且 $A\bar{A}$ 的任一负特征值至少必须有几何重数 2. 试与习题 5 比较.

7. 设 $A\in M_n$，且假定 λ 是 $A\bar{A}$ 的一个实负特征值，$A\bar{A}x=\lambda x$，$x\neq 0$，$\alpha^2=\lambda$，$A\bar{x}=\alpha y$，$A\bar{y}=\alpha x$. 根据习题 6，x 和 y 是无关的. (a)设 $x'=x+\beta y$，$y'=y-\bar{\beta}x$. 证明，对于 $\beta\in\mathbf{C}$ 的任一选择，$A\bar{x}'=\bar{\alpha}y'$且 $A\bar{y}'=\alpha x'$. (b)证明可选择 β 使 x' 与 y' 正交，并且选这样一个 β. (c)设 $s>0$ 使$\xi=sx'$是单位向量，又设 $\eta=sy'$. 证明，$A\bar{\xi}=\alpha\eta$，$A\bar{\eta}=\alpha\xi$ 和 $\xi^*\eta=0$. (d)设 $r>0$ 使 $r\eta$

[252] 是单位向量，又设 $U=[\eta\ r\eta\ u_3\cdots u_n]\in M_n$ 是酉矩阵. 证明

$$U^*A\bar{U}=\left[\begin{array}{cc|c} 0 & r\alpha & * \\ \bar{\alpha}/r & 0 & \\ \hline \multicolumn{2}{c|}{0} & A' \end{array}\right], \text{其中 } A'\in M_{n-2},$$

因而

$$U^*(A\bar{A})U = \begin{bmatrix} \lambda & 0 & * \\ 0 & \lambda & \\ \multicolumn{2}{c}{0} & A'\bar{A}' \end{bmatrix}.$$

(e)由此得出，$A\bar{A}$ 的每个负特征值有偶代数重数. 试与习题 6 比较.

8. 对任意 $A\in M_n$，证明

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}^{-1}\begin{bmatrix} A\bar{A} & 0 \\ \bar{A} & 0 \end{bmatrix}\begin{bmatrix} I & A \\ 0 & I \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \bar{A} & \bar{A}A \end{bmatrix}.$$

从这个明显的相似性推出，$A\bar{A}$ 和 $\bar{A}A$ 的有非零特征值的 Jordan 块之间存在一一对应. 因为 $\bar{A}A=\overline{A\bar{A}}$，证明 $A\bar{A}$ 的具有复特征值的 Jordan 块成共轭对出现. 由此推出，对任一 $A\in M_n$，$A\bar{A}$ 相似于实矩阵. **提示**：参看(3.4)中关于实 Jordan 形的讨论. 还有一些结论实际上是成立的. 事实上，$A\bar{A}$ 总相似于一个实矩阵的平方. 就 $A\bar{A}$ 的特征值而言，这意味着什么？

9. 如果 $A\in M_n$ 相似于实矩阵，证明 A 相似于 $\bar{A}$（反之亦然）. 利用这个事实和习题 9 证明，尽管 AB 一般不一定相似于 BA，但是，对任意 $A\in M_n$，$A\bar{A}$ 总相似于 $\bar{A}A$.

10. 说明 M_n 中可合对角化矩阵的集合包括以下集合：(a)只有实特征值的所有可对角化实矩阵. (b)具有 n 个线性无关的实特征向量的所有可对角化矩阵. (c)所有对称矩阵. (d)所有正定 Hermite 矩阵. **提示**：如果 A 是正定矩阵，则 $A=HH=H(HH^T)\bar{H}^{-1}$；H 是非奇异 Hermite 矩阵. (e)所有形如 AB 的矩阵，其中，A 是正定 Hermite 矩阵，而 B 是对称矩阵. 这与所有形如 H^2B 的矩阵的集合相同，其中，H 是非奇异 Hermite 矩阵，B 是对称矩阵. **提示**：$H^2B=H(HBH^T)\bar{H}^{-1}$.

253

11. 证明 M_n 中可合对角化矩阵的集合 CD_n 有下述性质：(a)如果 $A\in CD_n$，且 $S\in M_n$ 非奇异，则 $SA\bar{S}^{-1}\in CD_n$. (b)零矩阵在 CD_n 中. (c)如果 $A\in CD_n$ 且 $a\in\mathbf{C}$，则 $aA\in CD_n$. (d)如果 $A\in CD_n$ 可逆，则 $A^{-1}\in CD_n$.

12. 证明，(a)$\begin{bmatrix} i & 1 \\ 0 & i \end{bmatrix}$在普通意义下不能对角化，但它可合对角化. (b)$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$在普通意义下可对角化，但不能合对角化. (c)$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$既不能对角化也不能合对角化.

13. 如果 $A\in M_n$ 使得 $A\bar{A}=\Lambda=\lambda_1 I_{n_1}\oplus\cdots\oplus\lambda_k I_{n_k}$，其中，如果 $i\neq j$，则 $\lambda_i\neq\lambda_j$，且所有 $\lambda_i\geqslant 0$，证明，存在酉矩阵 $U\in M_n$，使得 $A=U\Delta U^T$，且 $\Delta=\Delta_1\oplus\cdots\oplus\Delta_k$，其中每个 $\Delta_i\in M_{n_i}$ 是上三角矩阵.

14. 引理(4.6.9)是说，对某个非奇异矩阵 $S\in M_n$，$A\in M_n$ 有分解 $A=S\bar{S}^{-1}$，当且仅当 $A\bar{A}=I$. 试用(4.4.4)证明，$A=U\bar{U}^{-1}=UU^T$ 对某个酉矩阵 $U\in M_n$ 成立，当且仅当 $A^{-1}=\bar{A}$ 且 A 是对称矩阵. 这与(4.4.7)有什么关系？

15. 设 $A\in M_n$，且记 $A=B+iC$，其中 B，$C\in M_n(\mathbf{R})$. 证明，$\lambda\in\mathbf{C}$ 是 A 的合特征向量，当且仅当 $\pm|\lambda|$ 是分块矩阵

$$F=\begin{bmatrix} B & C \\ C & -B \end{bmatrix}\in M_{2n}(\mathbf{R})$$

的(实)特征值. **提示**：用 $x=u+iv$ 表示 $A\bar{x}=rx$，u，$v\in\mathbf{R}^n$，$r=|\lambda|$. 因而，如果 F 没有

实特征值，则 A 不可能有合特征值.

16. 证明，如果 $A\in M_n$ 是对角矩阵或上三角矩阵，则 A 的特征值与 A 的合特征值在下述意义下是"相同"的：如果 λ 是 A 的特征值，则对所有 $\theta\in\mathbf{R}$，$e^{i\theta}\lambda$ 是 A 的合特征值，又如果 μ 是 A 的合特征值，则对某个 $\theta\in\mathbf{R}$，$e^{i\theta}\mu$ 是 A 的特征值.

17. 如果 $A\in M_n(\mathbf{R})$，证明，A 的每个实特征值也是 A 的合特征值. 又如果 $\mu\geqslant 0$ 是 A 的合特征值，则 μ 或 $-\mu$ 是 A 的特征值. **提示**：用 $x=u+iv$ 表示 $A\bar{x}=\mu x$，u，$v\in\mathbf{R}^n$. 考察例子 $A=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$ 以说明，一个实矩阵可以有这样的非实特征值，它不与任何合特征值相对应.

254

18. 当 $n=1$ 时引理(4.6.9)是什么意思？一个复数 z 位于复平面中的单位圆上是指 $\bar{z}z=1$. 这个条件到矩阵的普通推广是要求 $AA^*=I$；这样的矩阵称为酉矩阵，它们在矩阵理论中起着重要的作用. 另一个推广(当 $n=1$ 时它简化成同样的情形)是要求 $A\bar{A}=I$，而这些矩阵如引理(4.6.9)所描述的那样合相似于单位矩阵. 证明，若 $A\in M_n$ 且 $A\bar{A}=I$，则(a)A 是非奇异矩阵；(b)$A^{-1}=\bar{A}$；(c) $|\det A|=|\lambda_1\cdots\lambda_n|=1$；(d)若 $Ax=\lambda x$ 且 $x\neq 0$，则 $A\bar{x}=(1/\bar{\lambda})\bar{x}$；因而，只要 λ 是 A 的特征值，$1/\bar{\lambda}$ 就是 A 的特征值. 证明，若 $z\in\mathbf{R}$，$z\neq\pm 1$，则矩阵 $B=\begin{bmatrix}z & i\\ i & z\end{bmatrix}$ 有如下性质：$A=B\bar{B}^{-1}$ 的谱是

$$\left\{\frac{z-1}{z+1},\frac{z+1}{z-1}\right\},$$

因此，这样一些矩阵的特征值不都位于单位圆上.

19. 事实上，每个复矩阵 $A\in M_n$ 可以写成 $A=RE$，其中，R，$E\in M_n$，R 相似于实矩阵，而 $E\bar{E}=I$. 说明这个分解是如何从每个 $A\in M_n$ 相似于实矩阵的事实得来的，并且解释它是怎样推广了每个复数 z 可以写成 $z=re^{i\theta}$(其中 r 和 θ 是实数)这个事实.

20. 证明定理(4.6.11)可以由本节正文最后一段中所述的两个矩阵合相似的一般必要充分条件推出. **提示**：把条件应用于 A 和对角矩阵 Λ.

21. 利用每个 $A\in M_n$ 合相似于一个实矩阵的事实证明，如果 n 是奇数，则 A 至少必须有一个合特征值. **提示**：奇数阶实矩阵 R 至少有一个实特征值. 关于 R^2 的特征值，这意味着什么？如果 A 合相似于 R，$A\bar{A}$ 与 R^2 有何关系？

22. 设 $A\in M_n$ 是对称矩阵. 定理(4.6.11)后面的讨论说明 A 可对角化，所以，存在非奇异矩阵 $S\in M_n$ 和对角矩阵 $\Lambda\in M_n$，使得 $A=S\Lambda S^{-1}$. 说明，我们可以取 S 为酉矩阵[因而从定理(4.6.11)可导出推论(4.4.4)]，如下所述：注意到 A 的对称性推出 $(S^*S)\Lambda=\Lambda(\overline{S^*S})=\Lambda(S^*S)^T$. 利用极分解(7.3.3)把 S 写成 $S=UP$，其中，$U\in M_n$ 是酉矩阵，$P\in M_n$ 是 Hermite 矩阵，且对于某个多项式 $p(t)$，$P=p(S^*S)$[见定理(7.2.6)的证明]. 证明 $P\Lambda=\Lambda\bar{P}=\Lambda P^T$，因而 $S\Lambda S^{-1}=U\Lambda U^*$.

255

进一步阅读 关于合相似以及一个矩阵族同时合对角化的问题的更多信息，可参看(4.4)节和(4.5)节末所引用的 Hong 和 Horn 的文章，也可参看他们的报告："A Canonical Form for Matrices under Consimilarity"，*Linear Algebra Appl.* 102(1988)，143-168. 合相似的概念可作如下推广：用任意域代替复数域且用该域上的自同构代替复共轭运算：见[Jac]，p. 27.

256

第5章　向量范数和矩阵范数

5.0　导引

考虑 $\mathbf{C}^n$ 中的若干个向量或者 M_n 中的若干个矩阵，说有些向量或矩阵“小”，而说另一些向量或矩阵“大”，这是什么意思？在什么情况下可以说两个向量“很接近”或者“离得很远”？

在二维及三维实向量空间中，“大小”问题与“接近”问题通常涉及到 Euclid 距离．向量 $z\in\mathbf{R}^n$ 的 Euclid 长度是 $(z^Tz)^{1/2}=(\sum z_i^2)^{1/2}$，按该度量标准如果这个非负实数较小，就说 z 是“小向量”．此外，说向量 x 和 y 离得“很近”，是指差 $z=x-y$ 的 Euclid 长度是一个很小的非负实数．

矩阵可以看成高维空间中的向量，那么矩阵的“大小”指的是什么？关于无限维空间中的向量的“大小”指的是什么？关于复向量呢？除了用 Euclid 长度以外，还有没有度量实向量“大小”的其他有效方法？

要回答这些问题，一个办法是研究矩阵与向量的范数或大小的度量．范数可以看作 Euclid 长度的推广．当然，研究范数并不仅仅是为了作一下数学推广而已．为了恰当地表达象矩阵幂级数这样一些概念，范数是必不可少的，并且在分析和评价关于数值计算的各种算法中，它也是必需的．此外，已被采用的各种不同的范数大体上可适用于各种场合，所以，研究所有范数所共有的一些性质，而不是把注意力集中到任何个别的范数上，这样做是可取的．

257

下述各例从几个方面说明引进范数的必要性．

5.0.1 例(收敛性)　如果 x 是复数，且 $|x|<1$，我们知道，

$$(1-x)^{-1}=1+x+x^2+x^3+\cdots.$$

它给出计算方阵 $I-A$ 的逆的公式

$$(I-A)^{-1}=I+A+A^2+A^3+\cdots,$$

然而，什么时候这个公式才能成立呢？可以证明，这只要矩阵 A 的范数小于 1 就行了，并且这对于任何范数也是如此！同样，利用范数可以证明，许多其他可以用来定义一个矩阵的矩阵值函数的幂级数是收敛的，因而其定义是有意义的，例如，

$$e^A=\sum_{k=0}^{\infty}\frac{1}{k!}A^k$$

就是这样，为了按所要求的精确度计算一个特殊的函数值，范数也可以用来确定幂级数中所需要的项数，关于分析解方程组的迭代法的收敛性问题，也可以作类似的论述．

5.0.2 例(精确性)　设 f 是一个实变量的实值可微函数．我们知道，如果已知 $f(x)$ 在 $x=x_0$ 的值，则 f 在其附近 $x=x_0+h$ 的值可以用一阶导数

$$\frac{f(x_0+h)-f(x_0)}{h}=\frac{\Delta f}{\Delta x}\sim f'(x_0)$$

来估计．因此，在计算 f 与 x_0 的值时，如果实际上算出 f 在附近 $x=x_0+h$ 的值，那么，就有

一种估计相对误差的方法.

关于矩阵计算也会出现同样的问题. 假定想计算 A^{-1}(或 A 的其他函数), 但是 A 的各元是通过实验、通过其他数据的分析、或者是从先验的计算得来的, 并且不确切地知道这些数. 这时可以把 A 看成一个"真"的 A_0 再加上一个误差 E, 并且在计算 $A^{-1}=(A_0+E)^{-1}$ 而不是计算 A_0^{-1} 时, 我们想(用 E 的"大小")估计可能的"相对误差". 知道 A^{-1} 与 A_0^{-1} 之间的相差程度与知道 A^{-1} 的精确值可能是同样重要的, 而范数为处理这类问题提供了一个系统的方法. [258]

5.0.3 例(界) 对于与矩阵相关联的一些重要的量(例如特征值)的界常常要涉及范数. 当矩阵产生扰动时, 这些量的可能变化范围也要涉及范数.

5.1 向量范数和内积的定义性质

我们先考虑向量空间上的范数. 由于 M_n 是向量空间, 要做的一切也将适用于矩阵范数.

一个函数如果是范数, 它需要具备哪些性质. 为了说明这个问题, 对熟知的(实或复)纯量的绝对值概念进行抽象. 当然, 一个值得注意的差别是, 绝对值函数是实变量或复变量的实值函数, 而我们要求范数是描述一个向量的多元实值函数. $\mathbf{C}^n$ 上的这样一个函数是 Euclid 长度 $(z^*z)^{1/2}$. 但是还有另一些函数也具有 Euclid 长度的某些基本性质, 并且在某些场合可能是更为合适的度量, 它们可能提供另外的信息, 或者在某些内容中使用起来可能更为方便.

在整个这一章中, 只考虑实或复向量空间. 所有的主要结果对这两个域都成立, 不过在每一个结果中, 它必须与所采用的那个域相一致. 我们常用域 $\mathbf{F}$(一开始就视 $\mathbf{F}=\mathbf{R}$ 或 $\mathbf{C}$)叙述各个结果, 之后, 在余下的论述中涉及的是同一个域 $\mathbf{F}$.

5.1.1 定义 设 V 是域 $\mathbf{F}$($\mathbf{R}$ 或 $\mathbf{C}$)上的向量空间. 函数 $\|\cdot\|: V\to\mathbf{R}$ 称为向量范数, 是指对所有 $x, y\in V$ 有以下性质:

(1) $\|x\|\geqslant 0$, 非负性;

(1a) $\|x\|=0$ 当且仅当 $x=0$, 正定性;

(2) $\|cx\|=|c|\,\|x\|$ 对所有纯量 $c\in\mathbf{F}$ 成立, 齐次性;

(3) $\|x+y\|\leqslant\|x\|+\|y\|$, 三角不等式.

这四条公理是大家所熟悉的平面上的 Euclid 长度的几个性质. Euclid 长度还具有另外一些性质, 这些性质与这四条公理是独立无关的[例如, 平行四边形恒等式(5.1.8)], 因为它们对于一般的理论并不是必不可少的. 所以没有把它作为公理.

[259] 一个函数如果满足公理(1), (2), (3)但不一定满足(1a), 就称之为向量半范数. 半范数在允许某些非零向量有零长度方面推广了范数概念.

5.1.2 引理 如果 $\|\cdot\|$ 是 V 上的向量半范数, 则

$$|\,\|x\|-\|y\|\,|\leqslant\|x-y\|$$

对所有 $x, y\in V$ 成立.

证明: 因为 $y=x+(y-x)$, 由三角不等式和齐次性公理(2), 有

$$\|y\|\leqslant\|x\|+\|y-x\|=\|x\|+\|x-y\|.$$

由此可知

$$\|y\| - \|x\| \leqslant \|x-y\|.$$

又因为 $x=y+(x-y)$，所以再由三角不等式有

$$\|x\| \leqslant \|y\| + \|x-y\|,$$

因而

$$\|x\| - \|y\| \leqslant \|x-y\|.$$

因此我们证明了 $\pm(\|x\| - \|y\|) \leqslant \|x-y\|$，它等价于引理的论断. □

与 $\mathbf{C}^n$ 上 Euclid 长度相关的概念是通常的 Euclid 内积 y^*x(有时叫做"点积")，它有一些性质与两个向量间的"夹角"有关：如果 $y^*x=0$，那么 x 与 y 正交. 正象对向量范数所做的那样，可以抽象出 Euclid 内积的几个本质特征，并且用它们作为一般内积理论的公理.

5.1.3 定义 设 V 是域 $\mathbf{F}$($\mathbf{R}$ 或 $\mathbf{C}$)上的向量空间，函数 $\langle\cdot,\cdot\rangle: V\times V\to\mathbf{F}$ 是一个内积，是指对所有 $x, y, z\in V$，有

(1) $\langle x, x\rangle\geqslant 0$， 非负性；

(1a) $\langle x, x\rangle=0$ 当且仅当 $x=0$， 正定性；

(2) $\langle x+y, z\rangle=\langle x, z\rangle+\langle y, z\rangle$， 可加性；

(3) $\langle cx, y\rangle=c\langle x, y\rangle$ 对所有纯量 $c\in\mathbf{F}$ 成立， 齐次性；

(4) $\langle x, y\rangle=\overline{\langle y, x\rangle}$， Hermite 性.

练习 证明 Euclid 内积 $\langle x, y\rangle=y^*x$ 满足关于内积的所有上述四个公理. 260

练习 设 $D=\mathrm{diag}(d_1, d_2, \cdots, d_n)$，考虑函数 $\langle x, y\rangle\equiv y^*Dx$，$\langle\cdot,\cdot\rangle$ 满足关于内积的哪几条公理？D 在什么条件下 $\langle\cdot,\cdot\rangle$ 才是内积？

练习 由定义(5.1.3)的四条公理推导内积的下述性质：

(a) $\langle x, cy\rangle=\bar{c}\langle x, y\rangle$；

(b) $\langle x, y+z\rangle=\langle x, y\rangle+\langle x, z\rangle$；

(c) $\langle ax+by, cw+dz\rangle=a\bar{c}\langle x, w\rangle+b\bar{c}\langle y, w\rangle+a\bar{d}\langle x, z\rangle+b\bar{d}\langle y, z\rangle$；

(d) $\langle x, y\rangle=0$ 对所有 $y\in V$ 成立，当且仅当 $x=0$；

(e) $\langle x, \langle x, y\rangle y\rangle=|\langle x, y\rangle|^2$.

所有内积都共有的重要性质是 Cauchy-Schwarz 不等式.

5.1.4 定理(Cauchy-Schwarz 不等式) 如果 $\langle\cdot,\cdot\rangle$ 是域 $\mathbf{F}$($\mathbf{R}$ 或 $\mathbf{C}$)上的向量空间 V 上的内积，则

$$|\langle x,y\rangle|^2 \leqslant \langle x,x\rangle\langle y,y\rangle$$

对所有 $x, y\in\mathbf{F}$ 成立. 等式成立当且仅当 x 与 y 线性相关，即对某个 $\alpha\in\mathbf{F}$ 有 $x=\alpha y$ 或 $y=\alpha x$.

证明：设 $x, y\in V$ 是给定的. 如果 $y=0$，则结论显然成立，所以可以假定 $y\neq 0$. 设 $t\in\mathbf{R}$，考虑 $p(t)\equiv\langle x+ty, x+ty\rangle=\langle x, x\rangle+t\langle y, x\rangle+t\langle x, y\rangle+t^2\langle y, y\rangle=\langle x, x\rangle+2t\mathrm{Re}\langle x, y\rangle+t^2\langle y, y\rangle$，它是实系数二次多项式. 由公理(5.1.3(1))得知，对所有实数 t，$p(t)\geqslant 0$，因而 $p(t)$ 不能有实的单根. 因此，$p(t)$ 的判别式一定是非正的，即

$$(2\mathrm{Re}\langle x,y\rangle)^2 - 4\langle y,y\rangle\langle x,x\rangle \leqslant 0,$$

因而

$$(\mathrm{Re}\langle x,y\rangle)^2 \leqslant \langle x,x\rangle\langle y,y\rangle. \tag{5.1.5}$$

因为这个不等式对任一对向量都成立，所以当用$\langle x, y\rangle y$代替y时它也一定成立. 于是还有不等式

$$(\mathrm{Re}\langle x,\langle x,y\rangle y\rangle)^2 \leqslant \langle x,x\rangle\langle y,y\rangle \mid \langle x,y\rangle \mid^2.$$

但是，$\mathrm{Re}\langle x, \langle x, y\rangle y\rangle=\mathrm{Re}\overline{\langle x, y\rangle}\langle x, y\rangle=\mathrm{Re}\mid\langle x, y\rangle\mid^2=\mid\langle x, y\rangle\mid^2$，因此

$$\mid\langle x,y\rangle\mid^4 \leqslant \langle x,x\rangle\langle y,y\rangle \mid \langle x,y\rangle \mid^2. \tag{5.1.6}$$

[261] 如果$\langle x, y\rangle=0$，则定理的结论是显然的；否则可以用数$\mid\langle x, y\rangle\mid^2$除(5.1.6)两边得到所要证的不等式. 因为公理(1a)，所以仅当$x+ty=0$对某个t成立时$p(t)$才有一个实(二重)根. 因此，在判别条件(5.1.5)中等式能够成立，当且仅当x和y线性相关. □

5.1.7 推论 如果$\langle\cdot, \cdot\rangle$是$V$上的向量内积，则$\|x\|=(\langle x, x\rangle)^{1/2}$是$V$上的向量范数.

练习 证明(5.1.7). **提示**：只有三角不等式的验证不明显. 计算$\|x+y\|^2$，然后利用Cauchy-Schwarz不等式.

如果$\|\cdot\|$是适合$\|x\|=\langle x, x\rangle^{1/2}$的向量范数，其中$\langle\cdot, \cdot\rangle$为某个内积，就称该向量范数可由内积(即$\langle\cdot, \cdot\rangle$)诱导.

习题

1. 设e_i表示$\mathbf{C}^n$中的第i个单位坐标向量且假定$\|\cdot\|$是$\mathbf{C}^n$上的一个向量半范数，证明

$$\|x\| \leqslant \sum_{i=1}^n |x_i| \|e_i\|.$$

2. 如果$\|\cdot\|$是V上的一个向量半范数，证明$V_0=\{v\in V: \|v\|=0\}$是V的一个子空间(称为关于$\|\cdot\|$的零空间). (a)如果V_1是V的适合$V_0\cap V_1=\{0\}$的任一子空间，证明$\|\cdot\|$是V_1的一个向量范数. (b)考虑用

$$x\sim y \text{ 当且仅当 } \|x-y\|=0$$

定义的关系$x\sim y$. 证明，$\sim$是V上的一个等价关系，而这个等价关系的诸陪集具有形式$\hat{x}=\{x+y\in V: y\in V_0\}$，并且这些陪集的集合以一种自然的方式构成一个向量空间. 证明，函数$\|\hat{x}\|=\{\|x\|: x\in\hat{x}\}$是有意义的，并且是诸陪集组成的向量空间上的一个向量范数. (c)说明与每个向量半范数有关的自然范数为什么存在. (d)$\|x\|\equiv 0$是半范数吗？(e)给出一个不是范数的非平凡半范数的例子.

3. 证明，如果定义两个非零向量间的夹角是介于0与$\pi/2$之间的值

$$\cos^{-1}\left(\frac{|\langle x,y\rangle|}{(\langle x,x\rangle\langle y,y\rangle)^{1/2}}\right),$$

[262] 则这个夹角概念对任何内积都是有意义的.

4. 证明，如(5.1.7)中由内积诱导的向量范数一定满足平行四边形恒等式

$$\frac{1}{2}(\|x+y\|^2+\|x-y\|^2)=\|x\|^2+\|y\|^2. \tag{5.1.8}$$

为什么这个恒等式得此名称？事实上，等式(5.1.8)是一个给定的范数可由内积诱导的必要充分条件. 见习题10.

5. 考虑定义在 $\mathbf{C}^n$ 上的函数 $\|x\|_\infty \equiv \max_{1\leq i\leq n}|x_i|$. 证明 $\|\cdot\|_\infty$ 是不能由任何内积导出的向量范数.

6. 如果 $\|\cdot\|$ 是由内积 $\langle\cdot,\cdot\rangle$ 诱导的向量范数, 证明

$$\operatorname{Re}\langle x,y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2). \tag{5.1.9}$$

称这个等式为极化恒等式, 同时证明

$$\operatorname{Re}\langle x,y\rangle = \frac{1}{2}(\|x+y\|^2 - \|x\|^2 - \|y\|^2).$$

7. 证明, $\mathbf{C}^n$ 上的 l_1 范数 $\|x\|\equiv|x_1|+\cdots+|x_n|$ 适合公理(5.1.1), 但不适合极化等恒式(5.1.9). 因此, 它不是由任何内积导出的.

8. 如果 $\|\cdot\|$ 是由一个内积诱导的 V 上的向量范数, 则

$$\|x+y\|\,\|x-y\| \leqslant \|x\|^2+\|y\|^2$$

对所有 $x, y\in V$ 成立. 什么时候等式成立? 这个不等式对所有向量范数都成立吗? 给出这个不等式的几何解释.

9. 设 x 和 y 是 V 中给定的向量, V 有一个由内积 $\langle\cdot,\cdot\rangle$ 诱导的范数 $\|\cdot\|$, 并且假定 y 是非零的. 证明, 使 $\|x-\alpha y\|$ 的值达到极小的纯量 α_0 是 $\alpha_0=\langle x,y\rangle/\|y\|^2$, 并且 $x-\alpha_0 y$ 与 y 正交.

10. 不难证明(但也有一些技巧), 平行四边形恒等式(5.1.8)是一个给定的范数可由内积诱导的充分条件. 首先考虑 $\mathbf{R}$ 上的一个向量空间. 设 $\|\cdot\|$ 是 V 上一个给定的范数. (a)定义

$$\langle x,y\rangle = \frac{\|x+y\|^2 - \|x\|^2 - \|y\|^2}{2}. \tag{5.1.10}$$

证明, 用这种方法定义的 $\langle\cdot,\cdot\rangle$ 满足(5.1.3)中的公理(1), (1a)和(4), 且 $\langle x,x\rangle=\|x\|^2$. (b)利用(5.1.8)证明

$$\begin{aligned}4\langle x,y\rangle+4\langle z,y\rangle &= 2\|x+y\|^2+2\|z+y\|^2-2\|x\|^2-2\|z\|^2-4\|y\|^2\\ &= \|x+2y+z\|^2-\|x+z\|^2-4\|y\|^2 = 4\langle x+z,y\rangle,\end{aligned}$$

由此得出, 它也满足(5.1.3)中的可加性公理(2). (c)利用可加性公理证明, $\langle nx, y\rangle=n\langle x, y\rangle$ 且 $m\langle m^{-1}nx, y\rangle=\langle nx, y\rangle=n\langle x, y\rangle$, 其中 m 和 n 是非负整数. 利用(5.1.8)以及(5.1.10)证明 $\langle -x, y\rangle=-\langle x, y\rangle$. 由此得出 $\langle ax, y\rangle=a\langle x, y\rangle$, 其中 $a\in\mathbf{R}$ 是有理数. (d)设 $p(t)=t^2\|x\|^2+2t\langle x, y\rangle+\|y\|^2$, $t\in\mathbf{R}$, 然后证明, 若 t 为有理数, 则 $p(t)=\|tx+y\|^2$. 由 $p(t)$ 的连续性推出, 对所有 $t\in\mathbf{R}$, $p(t)\geqslant 0$. 根据 $p(t)$ 的判别式一定是非正的事实推出 Cauchy-Schwarz 不等式 $|\langle x, y\rangle|^2\leqslant\|x\|^2\|y\|^2$. (e)现在设 $a\in\mathbf{R}$ 是给定的, 证明

$$\begin{aligned}|\langle ax,y\rangle-a\langle x,y\rangle| &= |\langle(a-b)x,y\rangle+(b-a)\langle x,y\rangle|\\ &\leqslant|\langle(a-b)x,y\rangle|+|(b-a)\langle x,y\rangle|\leqslant 2|a-b|\,\|x\|\,\|y\|\end{aligned}$$

263

对任何有理数 b 成立, 并且可以使上界任意地小. 由此得出(5.1.10)满足(5.1.3)中的齐次性公理(3). 这就证明了 $\langle\cdot,\cdot\rangle$ 是 V 上的一个内积.

细心的读者会注意到, 在上述论证中关于范数 $\|\cdot\|$ 的三角不等式[(5.1.1)中的公理(3)]没有用到. 因此, (5.1.1)中的公理(1), (1a)和(2)连同(5.1.8)蕴涵以下事实: 函数 $\|\cdot\|$ 是由内

积诱导的，因而它是一个范数，并且一定满足三角不等式.（f）若 V 是一个复向量空间，定义

$$\langle x,y\rangle = \frac{\|x+y\|^2-\|x\|^2-\|y\|^2}{2}+\frac{i(\|x+iy\|^2-\|x\|^2-\|y\|^2)}{2}.$$

$\langle x, y\rangle$ 的实部是 V 看作 $\mathbf{R}$ 上一个向量空间时的内积. 利用这一事实和(5.1.8)证明 $\langle\cdot,\cdot\rangle$ 是 V 看作 $\mathbf{C}$ 上一个向量空间时的内积.

进一步阅读 平行四边形恒等式是一个给定的范数可由内积诱导的必要充分条件，这个结果的最早证明似乎应该属于 P. Jordan 和 J. Von Neumann，他们的文章是"On Inner Products in Linear Metric Spaces," *Ann. Math.* 36(2)(1935)，719—723. 习题 10 中所给出的这个结果的证明参考了下述文章：D. Fearnley-Sander and J. S. V. Symons，"Apollonius and Inner Products," *Amer. Math. Monthly* 81(1974)，990-993.

5.2 向量范数的例子

下面所列举的是一些常见的向量范数的例子.

5.2.1 $\mathbf{C}^n$ 上的 Euclid 范数(或 l_2 范数)是

$$\|x\|_2 \equiv (|x_1|^2+\cdots+|x_n|^2)^{1/2}.$$

这也许是大家最熟悉的向量范数，这是因为 $\|x-y\|_2$ 能度量两点 x，$y\in\mathbf{C}^n$ 之间的标准 Euclid 距离. 这个范数还是由普通的 Euclid 内积导出的；即 $\|x\|_2^2=\langle x, x\rangle=x^*x$.

264

练习 验证 $\|x\|_2$ 是 $\mathbf{C}^n$ 上的向量范数.

练习 范数 $\|\cdot\|$ 称为酉不变的，是指 $\|Ux\|=\|x\|$ 对所有 $x\in\mathbf{C}^n$ 和所有酉矩阵 $U\in M_n$ 成立. 证明 Euclid 范数 $\|\cdot\|$ 是酉不变的.

5.2.2 $\mathbf{C}^n$ 上的和范数(或 l_1 范数)是

$$\|x\|_1 \equiv |x_1|+\cdots+|x_n|.$$

因为其长度只是沿各坐标方向的直线度量，这个范数也称为 1—范数或更风趣地称为 Manhattan 范数.

练习 验证和范数是 $\mathbf{C}^n$ 上的向量范数，但不是由内积导出的. **提示**：利用(5.1.8).

5.2.3 $\mathbf{C}^n$ 上的极大范数(或 l_∞ 范数)是

$$\|x\|_\infty \equiv \max\{|x_1|,\cdots,|x_n|\}.$$

练习 证明 $\|\cdot\|_\infty$ 是 $\mathbf{C}^n$ 上的向量范数.

练习 $\|\cdot\|_\infty$ 是由内积导出的吗？

5.2.4 $\mathbf{C}^n$ 上的 l_p 范数是

$$\|x\|_p \equiv \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p}, p\geqslant 1.$$

练习 验证，对于 $p\geqslant1$，每个 l_p 范数是 $\mathbf{C}^n$ 上的向量范数，且对于每个 $x\in\mathbf{C}^n$，$\|x\|_\infty=\lim_{p\to\infty}\|x\|_p$. **提示**：只有三角不等式是较难验证的公理. 关于 l_p 范数的三角不等式是称之为 Minkowski 不等式的经典不等式.

练习 给出一个不是 l_p 范数的向量范数的例子.

上述各个向量范数的例子都是 $\mathbf{C}^n$ 上的范数，不过还能用它们来构造任意有限维实或复向量空间 V 上的向量范数，如果 $\mathscr{B}=\{b^{(1)},\cdots,b^{(n)}\}$ 是 V 的基，则 [265]

$$x\to[x]_{\mathscr{B}}\equiv\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\in\mathbf{C}^n,\ x=\sum_{i=1}^{n}x_ib^{(i)}$$

是 V 到 $\mathbf{C}^n$ 上的同构. 如果 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上的任一向量范数，则不难证明

$$\|x\|_{\mathscr{B}}\equiv\|[x]_{\mathscr{B}}\|=\|[x_1,\cdots,x_n]^T\|,\ x=\sum_{i=1}^{n}x_ib^{(i)}$$

是 V 上的向量范数.

练习 验证最后一个结论.

称矩阵 $B\in M_n$ 是关于 $\mathbf{C}^n$ 上的向量范数 $\|\cdot\|$ 的一个等距变换，是指

$$\|Bx\|=\|x\|$$

对所有 $x\in\mathbf{C}^n$ 成立.

练习 证明，关于任何向量范数的等距变换一定是非奇异矩阵.

练习 证明，关于给定的向量范数的各个变换等距所组成的矩阵集合构成一个群(称为该范数的等距变换群). 除了酉矩阵以外，对 $\|\cdot\|_2$，还存在任何等距变换吗?

练习 证明，和范数的等距变换群是所有那样一些矩阵组成的集合(群)，这些矩阵看上去像置换矩阵，只是其中的元素"+1"用绝对值为 1 的任意复数来代替.

练习 极大范数的等距变换群是什么?

向量范数的定义并未要求向量空间 V 是有限维的. 例如，空间 V 可能是实区间$[a,b]$上的所有实或复值连续函数构成的向量空间 $C[a,b]$.

5.2.5 例 $C[a,b]$上的某些范数例子类似于已经对 $\mathbf{C}^n$ 定义的诸范数. 例如，

$$\|f\|_2\equiv\left[\int_a^b|f(t)|^2\mathrm{d}t\right]^{1/2},\qquad L_2\text{ 范数；}$$

$$\|f\|_1\equiv\int_a^b|f(t)|\mathrm{d}t,\qquad L_1\text{ 范数；}$$ [266]

$$\|f\|_p\equiv\left[\int_a^b|f(t)|^p\mathrm{d}t\right]^{1/p},\ p\geqslant1,\qquad L_p\text{ 范数；}$$

$$\|f\|_\infty=\max\{|f(x)|:x\in[a,b]\},\qquad L_\infty\text{ 范数；}$$

它们都是 $C[a,b]$上的范数.

习题

1. 证明，如果 $0<p<1$，则(5.2.4)定义 $\mathbf{C}^n$ 上的一个函数，除了向量范数的一条公理以外，这个函数满足所有的公理. 它不满足哪条公理? 试给一个例子.

2. 设 $f\in C[0,1]$. 证明 $\|f\|_\infty=\lim\limits_{p\to\infty}\|f\|_p$.

3. 关于 $C[0,1]$上的 $\|\cdot\|_p$，其三角不等式看来与什么一样? 对于 $\mathbf{C}^n$，如何从 Minkowski 不等式(附录 B)出发去证明这个三角不等式.

4. 设 p_1，p_2，…，p_n 是给定的正实数. 下面哪一个是 $\mathbf{C}^n$ 上的向量范数？

(a) $\|x\|=\sum_{i=1}^{n}p_i|x_i|$，

(b) $\|x\|=\left(\sum_{i=1}^{n}p_i|x_i|^2\right)^{1/2}$，

(c) $\|x\|=\max\{p_1|x_1|,\cdots,p_n|x_n|\}$.

5. 设 $x_0\in[a,b]$是给定的点. 证明，如果 $a<b$，则 $C[a,b]$上的函数 $\|f\|_{x_0}\equiv|f(x_0)|$ 是半范数而不是范数.

6. 如果 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上的酉不变向量范数，证明，$\|\cdot\|=\alpha\|\cdot\|_2$ 对某个 $\alpha>0$ 成立，且 $\|\cdot\|_2$ 是使 $\|e_1\|=1$ 的唯一酉不变向量范数.

7. 证明，$\|y\|_\infty=\max_{\|x\|_1=1}|y^*x|$ 且 $\|x\|_1=\max_{\|y\|_\infty=1}|x^*y|$.

8. 试用前面的练习证明，如果 A^* 在和范数的等距变换群中，则 A 在极大范数的等距变换群中，反之亦然.

9. 所有 L_p 范数的所有等距变换群的交是什么？

267 **进一步阅读** 关于 Minkowski 和 Hölder 的经典不等式的详细讨论可参看[BB].

5.3 向量范数的代数性质

从任意一个或几个给定的范数出发可以用多种方式来构造一些新的范数. 例如，容易证明，两个向量(半)范数的和是一个向量(半)范数，一个向量(半)范数的任意正数倍又是一个(半)范数. 按另一种方式也不难证明，如果 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是向量范数，则由 $\|x\|\equiv\max\{\|x\|_\alpha,\|x\|_\beta\}$ 定义的函数 $\|\cdot\|$ 也是向量范数. 这些论断都是下述结果的特殊情形.

5.3.1 定理 如果 $\|\cdot\|_{\alpha_1}$，…，$\|\cdot\|_{\alpha_m}$ 是域 $\mathbf{F}$($\mathbf{R}$ 或 $\mathbf{C}$)上的向量空间 V 上的向量范数，又设 $\|\cdot\|_\beta$ 是 $\mathbf{R}^m$ 上的向量范数，使得对具有非负分量的所有 $y,z\in\mathbf{R}^m$ 都有 $\|y\|_\beta\leqslant\|y+z\|_\beta$. 则由 $\|x\|\equiv\|[\|x\|_{\alpha_1},\cdots,\|x\|_{\alpha_m}]^T\|_\beta$ 定义的函数 $\|\cdot\|:V\to\mathbf{R}$ 是 V 上的向量范数.

定理中关于范数 $\|\cdot\|_\beta$的单调性假设可确保所构造的函数 $\|\cdot\|$ 满足三角不等式. 所有的 l_p 范数都具有单调性这一性质，$\mathbf{R}^m$ 上的任一向量范数 $\|x\|_\beta$倘若只是 x 的分量的绝对值的函数，该范数也具有单调性这一性质；见(5.5.9 和 5.5.10). 但是，有一些向量范数不具有这一性质.

练习 证明定理(5.3.1).

练习 证明两个向量范数的和或取极大仍是向量范数的事实是定理(5.3.1)的特殊情形. 关于取极小又如何？

练习 设 $m=2$，$V=\mathbf{R}^2$ 且 $\|x\|_\beta=|x_1-x_2|+|x_2|$. 证明，$\|\cdot\|_\beta$ 是 $\mathbf{R}^2$ 上的向量范数，但函数 $\|x\|\equiv\|[\|x\|_\infty,\|x\|_1]^T\|_\beta=\min\{|x_1|,|x_2|\}+|x_1|+|x_2|$ 不是向量范数. $\|\cdot\|$ 满足哪几个向量范数公理？**提示**：考察 $x=\begin{bmatrix}1\\0\end{bmatrix}$，$y=\begin{bmatrix}0\\1\end{bmatrix}$，$\|x+y\|$，$\|x\|$ 以及 $\|y\|$. 这与(5.3.1)矛盾吗？

从旧范数构造新范数的另一种方式是由下面的结果给出的.

5.3.2 定理 如果 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上的向量范数，又如果 $T\in M_n$ 是非奇异矩阵，则由

$$\|x\|_T = \|Tx\|, \quad x\in \mathbf{C}^n$$

定义的 $\|\cdot\|_T$ 也是 $\mathbf{C}^n$ 上的向量范数.

268

练习 证明(5.3.2).

练习 在(5.3.2)中如果 T 是奇异矩阵，结果有何变化?

练习 为什么 $\|x\|=(|2x_1-3x_2|^2+|x_2|^2)^{1/2}$ 一定是 $\mathbf{C}^2$ 上的范数.（请不用计算!）

利用对偶的概念可以从旧范数构造新范数. 这种方法在(5.4)节末讨论.

习题

1. 如果 $\|\cdot\|$ 是向量半范数. 证明，对任意 $T\in M_n$，$\|x\|_T\equiv\|Tx\|$ 也是向量半范数. 如果 $\|\cdot\|$ 是向量范数，则 $\|\cdot\|_T$ 是向量半范数，它的零空间就是 T 的零空间.

2. 证明，任何向量半范数都具有形式 $\|\cdot\|_T$ 其中，$\|\cdot\|$ 是某个向量范数，而 $T\in M_n$ 是某个矩阵.

5.4 向量范数的分析性质

前两节中的各个例子清楚地说明，有许多各不相同的函数 $\|\cdot\|: V\to\mathbf{R}$ 满足范数的各条公理. 这样，有多种不同的范数可供选用. 这是很有用的，因为对于某种应用来说，一种范数可能比另一种范数更方便或更合适. 例如，l_2 范数用于优化问题往往是方便的，因为(除了原点以外)它是连续可微的. 另一方面，l_1 范数虽然只在一个较小的集合上可微，但是它在统计学中较为通用，因为它(是)导出的估计可能比经典的回归估计更有说服力. l_∞ 范数往往用起来是最自然的范数，因为它直接检验各个分量的收敛性，然而，遗憾的是，不论在分析上还是在代数上，它使用起来可能都不方便. 在实际应用中，基于一种范数可以很自然地建立起一种理论，与在某种场合很容易计算这种范数可能不是一回事. 因此，知道两种不同的范数之间可能存在什么关系是很重要的. 幸运的是，在有限维情形，所有范数都在某种很强的意义下是"等价"的.

分析中的基本概念是*序列的收敛*概念，而向量范数可以用来度量向量序列的收敛性.

5.4.1 定义 设 V 是 $\mathbf{R}$ 或 $\mathbf{C}$ 上的向量空间，且 $\|\cdot\|$ 是 V 上的范数，说 V 中的向量序列 $\{x^{(k)}\}$ 关于范数 $\|\cdot\|$ 收敛于 $x\in V$，当且仅当 $k\to\infty$ 时 $\|x^{(k)}-x\|\to 0$.

269

如果 $\{x^{(k)}\}$ 关于范数 $\|\cdot\|$ 收敛于 x，则记

$$x^{(k)}\xrightarrow[\|\cdot\|]{} x, \text{或关于 } \|\cdot\|, \lim_{k\to\infty} x^{(k)} = x.$$

应当弄清楚的是，上面提到的收敛性与哪一种范数有关，于是，关于一个给定的向量序列是否可能对一种范数收敛而对另一种范数则不收敛的问题提出来了. 这种二重性在无限维空间中可能出现.

5.4.2 例 考虑 $C[0,1]$($[0,1]$上的所有实值或复值连续函数组成的空间)中的函数序列 $\{f_k\}$，定义如下：对 $k=2,3,4,\cdots$，

$$f_k(x)=0, \qquad 0\leqslant x\leqslant \frac{1}{k};$$

$$f_k(x)=2(k^{5/2}x-k^{3/2}), \qquad \frac{1}{k}\leqslant x\leqslant \frac{3}{2k};$$

$$f_k(x)=2(-k^{3/2}x+2k^{1/2}),\qquad \frac{3}{2k}\leqslant x\leqslant\frac{2}{k};$$

$$f_k(x)=0,\qquad \frac{2}{k}\leqslant x\leqslant 1.$$

于是可以算出，

当 $k\to\infty$ 时，$\|f_k\|_1=\frac{1}{2}k^{-1/2}\to 0$；

对所有 k，$\|f_k\|_2=\frac{1}{\sqrt{3}}$；

当 $k\to\infty$ 时，$\|f_k\|_\infty=k^{1/2}\to\infty$.

所以，关于 L_1，$\lim\limits_{k\to\infty}f_k=0$，但是关于其他两种范数就不为零.

练习 画出上例中所描述的各个函数的草图，并且验证关于 L_1，L_2 和 L_∞ 范数所做的论断.

练习 设 $\|\cdot\|$ 是向量范数，如果 $x^{(k)}\xrightarrow[\|\cdot\|]{}x$，且 $x^{(k)}\xrightarrow[\|\cdot\|]{}y$ 利用三角不等式证明 $x=y$. 因此，关于给定的范数，讨论一个序列的极限(如果有的话)是有意义的.

所幸的是，例(5.4.2)中出现的现象在有限维向量空间的情形不可能发生. 为了看出这一点，需要一个关于范数连续性的一般引理.

270

5.4.3 引理 设 $\|\cdot\|$ 是域 **F**(**R** 或 **C**)上的向量空间 V 上的范数，又设 $x^{(1)}$，$x^{(2)}\cdots$，$x^{(m)}\in V$ 是给定的向量. 则用

$$g(z_1,z_2,\cdots,z_m)\equiv\|z_1x^{(1)}+z_2x^{(2)}+\cdots+z_mx^{(m)}\|$$

定义的函数 g：$\mathbf{F}^m\to\mathbf{R}$ 是一致连续的函数.

证明：设 $u=\sum\limits_{i=1}^m u_ix^{(i)}$ 和 $v=\sum\limits_{i=1}^m v_ix^{(i)}$，经计算，

$$|g(u_1,\cdots,u_m)-g(v_1,\cdots,v_m)|=|\,\|u\|-\|v\|\,|\leqslant\|u-v\|$$
$$=\left\|\sum_{i=1}^m(u_i-v_i)x^{(i)}\right\|\leqslant\sum_{i=1}^m|u_i-v_i|\,\|x^{(i)}\|\leqslant C\max_{1\leqslant i\leqslant m}|u_i-v_i|,$$

其中 $C\equiv m\max\limits_{1\leqslant i\leqslant m}\|x^{(i)}\|$. 第一个不等式可由引理(5.1.2)得到. 注意，有限常数 C 只依赖于范数 $\|\cdot\|$ 与 m 个向量 $x^{(1)}$，…，$x^{(m)}$. 如果向量 $x^{(i)}$ 都是零向量，则无需证明，否则有 $C>0$. 为了使 $|g(u_1,\cdots,u_m)-g(v_1,\cdots,v_m)|<\varepsilon$，我们只需选取 $|u_i-v_i|<\varepsilon/C$. □

虽然引理并未要求向量空间是有限维的，但是，向量 $x^{(i)}$ 的个数有限是很重要的.

练习 试用这个引理推出，$\mathbf{R}^n$ 或 $\mathbf{C}^n$ 上的每个向量范数都是一致连续函数.

然而，V_n 的有限维性质对下面的基本事实是必不可少的.

5.4.4 定理 设 f_1 和 f_2 是域 **F**(**R** 或 **C**)上的有限维向量空间 V 上的两个实值函数，又设 $\mathscr{B}=\{x^{(1)},\cdots x^{(n)}\}$ 是 V 的基. 假定 f_1 和 f_2 是

(a) 正定的：对所有 $x\in V$ 有 $f_i(x)\geqslant 0$，$f_i(x)=0$ 当且仅当 $x=0$；

(b) 齐次的：对所有 $\alpha\in\mathbf{F}$，$x\in V$，$f_i(\alpha x)=|\alpha|f_i(x)$；

(c) 连续的：$f_i(x(z))$ 在 $\mathbf{F}^n$ 上连续，其中

$$z=[z_1,\cdots,z_n]^T\in\mathbf{F}^n\quad 且\quad x(z)\equiv z_1x^{(1)}+\cdots+z_nx^{(n)}.$$

则存在有限正常数 C_m 和 C_M，使得

$$C_m f_1(x) \leqslant f_2(x) \leqslant C_M f_1(x)$$

对所有 $x\in V$ 成立.

证明：我们在 Euclid 单位球面 $S=\{z\in \mathbf{F}^n: \|z\|_2=1\}$（它是 $\mathbf{F}^n$ 中的紧集）上定义 $h(z)\equiv f_2(x(z))/f_1(x(z))$. 由(a)可知，$h(z)$的分母在 S 上不为零，因而根据(c)可知 $h(z)$在 S 上连续. 由 Weierstrass 定理(见附录 E)，连续函数 h 在紧集 S 上达到一个有限的正极大值 C_M 和一个正极小值 C_m，因而 271

$$C_m f_1(x(z)) \leqslant f_2(x(z)) \leqslant C_M f_1(x(z))$$

对所有 $z\in S$ 成立. 由于 $z/\|z\|_2\in S$ 对每个非零 $z\in \mathbf{F}^n$ 成立，(b)保证这两个不等式对所有非零 $z\in \mathbf{F}^n$ 成立；因为 $f_i(0)=0$，所以它们对 $z=0$ 显然成立. 但每个 $x\in V$ 都具有形式 $x=x(z)$，其中 $z\in \mathbf{F}^n$ 为某个向量，这是因为 $\mathscr{B}$ 是基，因此要证的不等式对所有 $x\in V$ 成立. □

定义 设 V 是实或复向量空间. 函数 f: $V\to\mathbf{R}$ 如果满足定理(5.4.4)中的正定性，齐次性和连续性等三个假设条件，则称这个函数是*准范数*.

当然，一类准范数的最重要的例子是向量范数；引理(5.4.3)说明，每个向量范数满足定理(5.4.4)的连续性假设(c). 一个满足三角不等式的准范数是向量范数. 由于这类范数的重要性，我们把这种情形下所得到的结果叙述成下面的推论.

5.4.5 推论 设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是有限维实或复向量空间 V 上的任意两个向量范数. 则存在有限正常数 C_m 和 C_M 使得 $C_m\|x\|_\alpha\leqslant\|x\|_\beta\leqslant C_M\|x\|_\alpha$ 对所有 $x\in V$ 成立.

练习 (5.4.5)为何对向量半范数不成立?

练习 设 $x=[x_1, x_2]^T\in\mathbf{R}^2$，且考虑 $\mathbf{R}^2$ 上的下列两种范数：$\|x\|_\alpha\equiv\|[10x_1, x_2]^T\|_\infty$ 和 $\|x\|_\beta\equiv\|[x_1, 10x_2]^T\|_\infty$. 证明函数 $f(x)\equiv(\|x\|_\alpha\|x\|_\beta)^{1/2}$ 是 $\mathbf{R}^2$ 上的准范数而不是范数. **提示**：考虑 $f([1, 1]^T)$，$f([0, 1]^T)$和 $f([1, 0]^T)$.

练习 如果 $\|\cdot\|_{\alpha_1}$，…，$\|\cdot\|_{\alpha_k}$ 是 V 上的向量范数，证明，$f(x)\equiv(\|x\|_{\alpha_1}\cdots\|x\|_{\alpha_k})^{1/k}$ 和 $h(x)\equiv\min\{\|x\|_{\alpha_1}, \cdots, \|x\|_{\alpha_k}\}$ 是 V 上的准范数，但不一定是向量范数.

(5.4.5)的一个推论是，有限维复向量空间中的一个向量序列(关于范数)的收敛性与所采用的范数无关.

5.4.6 推论 如果 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是有限维实或复向量空间的向量范数，又如果$\{x^{(k)}\}$是给定的向量序列，则关于 $\|\cdot\|_\alpha$ 有 $\lim\limits_{k\to\infty}x^{(k)}=x$，当且仅当关于 $\|\cdot\|_\beta$ 有 $\lim\limits_{k\to\infty}x^{(k)}=x$. 272

证明：因为 $C_m\|x^{(k)}-x\|_\alpha\leqslant\|x^{(k)}-x\|_\beta\leqslant C_M\|x^{(k)}-x\|_\alpha$ 对所有 k 成立. 由此可知，当 $k\to\infty$时，$\|x^{(k)}-x\|_\alpha\to0$ 当且仅当 $\|x^{(k)}-x\|_\beta\to0$.

5.4.7 定义 两个范数称为是等价的，是指只要序列$\{x_k\}$对第一个范数收敛于向量 x，则它对第二个范数收敛于同一向量. 因此(5.4.6)是说，*对有限维实或复向量空间，所有范数都是等价的*. 在例(5.4.2)中已经看到，在无限维空间中两个范数可以不等价.

由于 $\mathbf{R}^n$ 或 $\mathbf{C}^n$ 上的所有范数都与 $\|\cdot\|_\infty$ 等价，所以对任何向量范数有 $\lim\limits_{k\to\infty}x^{(k)}=x$，当且仅当对所有 $i=1, 2, \cdots, n$ 有

$$\lim_{k\to\infty} x_i^{(k)} = x_i.$$

这就是说，(关于任一个基)依分量收敛等价于对任何向量范数收敛.

在有限维情形，关于所有向量范数都等价的另一重要推论是，每个向量范数的单位球及单位球面是紧的. 这个事实蕴涵任意向量范数的单位球上的连续复值函数是有界的，并且，如果这个函数是实的，则它达到它的极大值和极小值.

5.4.8 推论 设 V 是 $\mathbf{R}^n$ 或 $\mathbf{C}^n$. 设 $f(\cdot)$是 V 上的准范数. 则集合$\{x: f(x)\leqslant 1\}$和$\{x: f(x)=1\}$是紧的. 特别是，如果$\|\cdot\|$是 V 上的向量范数，则闭单位球$\{x: \|x\|\leqslant 1\}$及单位球面$\{x: \|x\|=1\}$都是紧的.

证明：由(5.4.4)，存在某个 $C>0$，使得$\|x\|_2\leqslant Cf(x)$对所有 $x\in V$ 成立，所以集合$\{x: f(x)\leqslant 1\}$是有界集，它包含在一个半径为 C，中心在原点的普通 Euclid 球中. 因为 $f(\cdot)$连续，所以集合$\{x: f(x)=1\}$与$\{x: f(x)\leqslant 1\}$都是闭的. 由于 $\mathbf{R}^n$ 或 $\mathbf{C}^n$ 中的有界闭集是紧的，我们完成了证明. □

我们遇到的问题常常不是判定给定的序列$\{x^{(k)}\}$是否收敛于某个给定向量 x，而是判别给定的序列$\{x^{(k)}\}$是否收敛. 因为这个缘故，需要有一个收敛性判别准则，这个准则与序列所收敛的极限 x 没有关系. 如果存在这样一个极限 x，那么，当 k、$j\to\infty$时，

273

$$\|x^{(k)}-x^{(j)}\| = \|x^{(k)}-x+x-x^{(j)}\| \leqslant \|x^{(k)}-x\| + \|x-x^{(j)}\| \to 0.$$

这是促成给出下述定义的原因.

5.4.9 定义 向量空间 V 中的序列$\{x^{(k)}\}$是关于范数$\|\cdot\|$的 Cauchy 序列，指的是，对任意 $\varepsilon>0$，都存在正整数 $N(\varepsilon)$，只要 k_1，$k_2\geqslant N(\varepsilon)$，就有

$$\|x^{(k_1)}-x^{(k_2)}\|\leqslant\varepsilon.$$

5.4.10 定理 设$\|\cdot\|$是有限维实或复向量空间 V 上给定的范数，$\{x^{(k)}\}$是 V 中给定的向量序列. 则序列$\{x^{(k)}\}$收敛于 V 中的一个向量，当且仅当它是关于范数$\|\cdot\|$的 Cauchy 序列.

证明：选取 V 的基 $\mathscr{B}$，且考虑等价的范数$\|[x]_{\mathscr{B}}\|_\infty$，注意到，如果假定 $V=\mathbf{R}^n$ 或 $\mathbf{C}^n$(n 为某个整数)且假定其范数是$\|\cdot\|_\infty$，也不会影响证明的一般性，如果$\{x^{(k)}\}$是 Cauchy 序列，则对每个 $i=1, 2, \cdots$，实数或复数的每个分量序列$\{x_i^{(k)}\}$也是 Cauchy 序列. 由于实数或复数的 Cauchy 序列一定有极限，因而，对每个 $i=1, 2, \cdots, n$，都存在一个纯量 x_i，使得$\lim\limits_{k\to\infty} x_i^{(k)}=x_i$；容易验证$\lim\limits_{k\to\infty} x^{(k)}=x$，这里 $x=[x_1, \cdots, x_n]^T$，另一方面，如果存在 x 使$\lim\limits_{k\to\infty} x^{(k)}=x$，则$\|x^{(k_1)}-x^{(k_2)}\|\leqslant\|x^{(k_1)}-x\|+\|x-x^{(k_2)}\|$，因而给定的序列是 Cauchy 序列. □

(用于上述定理证明中的)实数域或复数域的一个基本性质是，一个序列是 Cauchy 序列，当且仅当它收敛于某个(实或复)纯量. 这个性质叫做实数域和复数域的完备性. 刚才证明了，对于任何向量范数，完备性都可以推广到有限维实和复向量空间，遗憾的是，完备性对非有限维的向量空间不一定成立.

5.4.11 定义 具有范数$\|\cdot\|$的向量空间 V 称为对范数$\|\cdot\|$是完备的，指的是关于该范数$\|\cdot\|$是 Cauchy 序列的每个序列都收敛于 V 的一点.

274

练习 考虑具有 L_1 范数 $\|f\|_1=\int_0^1 |f(t)|\,\mathrm{d}t$ 的向量空间，并且考虑由

$$f_k(t)=0, \qquad 0\leqslant t\leqslant \frac{1}{2}-\frac{1}{k};$$

$$f_k(t)=\frac{k}{2}\left(t-\frac{1}{2}+\frac{1}{k}\right), \qquad \frac{1}{2}-\frac{1}{k}\leqslant t\leqslant \frac{1}{2}+\frac{1}{k};$$

$$f_k(t)=1, \qquad \frac{1}{2}+\frac{1}{k}\leqslant t\leqslant 1$$

定义的函数．画出各函数 f_k 的草图．证明，$\{f_k\}$ 是 Cauchy 序列，但是不存在函数 $f\in C[0,1]$，使得对范数 $\|\cdot\|_1$，有 $\lim_{k\to\infty} f_k=f$.

利用 $\mathbf{R}^n$ 或 $\mathbf{C}^n$ 上的任意范数或准范数的单位球是紧集的事实，可以引进另一种由旧范数生成新范数的有效方法．

5.4.12 定义 设 $f(\cdot)$ 是 $V=\mathbf{R}^n$ 或 $\mathbf{C}^n$ 上的准范数．函数

$$f^D(y)\equiv \max_{f(x)=1}\operatorname{Re} y^*x$$

称为 f 的对偶范数．

首先说明，对偶范数是 V 上有意义的函数，这是因为对于每个固定的 $y\in V$，$\operatorname{Re} y^*x$ 是 x 的连续函数．又由(5.4.8)，集合 $\{x: f(x)=1\}$ 是紧集．根据 Weierstrass 定理，$\operatorname{Re} y^*x$ 的极大值在某点 $x_0\in\{x: f(x)=1\}$ 达到．如果 c 是纯量，且 $|c|=1$，则由 f 的齐次性有

$$\max_{f(x)=1}|y^*x|=\max_{f(x)=1}\max_{|c|=1}\operatorname{Re} cy^*x=\max_{f(x)=1}\max_{|c|=1}\operatorname{Re} y^*(cx)$$

$$=\max_{|c|=1}\max_{f(cx)=1}\operatorname{Re} y^*x=\max_{f(x)=1}\operatorname{Re} y^*x,$$

所以，关于对偶范数，有一个有时是简便的等价定义，那就是

$$f^D(y)=\max_{f(x)=1}|y^*x|. \tag{5.4.12a}$$

最后，必须说明，关于函数 f^D 的对偶范数名称是名符其实的．函数 $f^D(\cdot)$ 显然是齐次的和正定的，因为，如果 $y\neq 0$，则可以利用 $f(\cdot)$ 的齐次性证明

$$f^D(y)=\max_{f(x)=1}|y^*x|\geqslant\left|y^*\frac{y}{f(y)}\right|=\frac{\|y\|_2^2}{f(y)}>0.$$

275

也许值得注意的是，即使函数 $f(\cdot)$ 不满足三角不等式，但是它的对偶 $f^D(\cdot)$ 仍恒满足

$$f^D(y+z)=\max_{f(x)=1}|(y+z)^*x|\leqslant\max_{f(x)=1}[|y^*x|+|z^*x|]$$

$$\leqslant\max_{f(x)=1}|y^*x|+\max_{f(x)=1}|z^*x|=f^D(y)+f^D(z).$$

因此，一个准范数的对偶范数总是范数．

这样，采用构造其对偶范数的方法，任一准范数可生成一个范数．这种构造法常常用于准范数实际上是范数的情形．

关于对偶范数的一个简单不等式在下面的引理中给出．我们将会看到，它是 Cauchy-Schwarz 不等式的自然推广．

5.4.13 引理 设 $f(\cdot)$ 是 $V=\mathbf{R}^n$ 或 $\mathbf{C}^n$ 上的准范数．则

$$|y^*x| \leqslant f(x)f^D(y),$$
$$|y^*x| \leqslant f^D(x)f(y)$$

对所有 x, $y \in V$ 成立.

证明: 如果 $x \neq 0$, 则

$$\left|y^* \frac{x}{f(x)}\right| \leqslant \max_{f(z)=1} |y^*z| = f^D(y).$$

因而 $|y^*x| \leqslant f(x)f^D(y)$. 由于这个不等式对 $x=0$ 也成立, 所以证明了第一个不等式. 因为 $|y^*x| = |x^*y|$, 第二个不等式可由第一个不等式得到. □

要识别一些最常见的向量范数的对偶范数并不困难. 如果 x, $y \in \mathbf{C}^n$, 则 Hölder 不等式的特殊情形是

$$|y^*x| = \left|\sum_{i=1}^n \bar{y}_i x_i\right| \leqslant \sum_{i=1}^n |\bar{y}_i x_i| \leqslant \max_{1 \leqslant i \leqslant n} |y_i| \sum_{j=1}^n |x_j| = \|y\|_\infty \|x\|_1. \quad (5.4.14)$$

如果 y 是给定的向量, 又如果 x 是(关于 $\|\cdot\|_1$ 的)单位向量, 且对使 $|y_i| = \|y\|_\infty$ 的某个 i, 有 $x_i=1$, 否则有 $x_i=0$, 那么(5.4.14)中的等式成立. 同理, 如果 x 是给定的非零向量, 又如果 y 是(关于 $\|\cdot\|_\infty$ 的)单位向量, 且对所有使 $x_i \neq 0$ 的 i 有 $y_i = x_i/|x_i|$, 否则有 $y_i=0$. 则(5.4.4)中的等式成立. 因此,

$$(\|y\|_1)^D = \max_{\|x\|_1=1} |y^*x| = \max_{\|x\|_1=1} \|y\|_\infty \|x\|_1 = \|y\|_\infty,$$
$$(\|y\|_\infty)^D = \max_{\|x\|_\infty=1} |y^*x| = \max_{\|x\|_\infty=1} \|y\|_1 \|x\|_\infty = \|y\|_1.$$

276 由此得出$(\|\cdot\|_1)^D = \|\cdot\|_\infty$ 和$(\|\cdot\|_\infty)^D = \|\cdot\|_1$.

如果考虑 Euclid 范数 $\|\cdot\|_2$, 给定的非零向量 y 和任意向量 x, 则 Cauchy-Schwarz 不等式是指,

$$|y^*x| = \left|\sum_{i=1}^n \bar{y}_i x_i\right| \leqslant \|y\|_2 \|x\|_2, \quad (5.4.15)$$

且当 $x = y/\|y\|_2$ 时等式成立. 运用上述关于 l_1 和 l_∞ 范数的同样的论证, 得知$(\|y\|_2)^D = \|y\|_2$, 所以 Euclid 范数是它自身的对偶范数.

练习 说明为什么(5.4.13)中的不等式是 Cauchy-Schwarz 不等式(5.1.4)的推广.

注意, 对刚才讨论的三个范数(l_1, l_2 和 l_∞)中的每一个, 其对偶范数的对偶是原范数. 这不是偶然的巧合: 对偶性定理(5.5.14)说明, 这种情形总是成立的.

5.4.16 定理 设 $\|\cdot\|$ 是 $V=\mathbf{R}^n$ 或 $\mathbf{C}^n$ 上的向量范数, $\|\cdot\|^D$ 是它的对偶范数, 又设 $c>0$ 是给定的. 则 $\|x\| = c\|x\|^D$ 对所有 $x \in V$ 成立当且仅当 $\|\cdot\| = \sqrt{c}\|\cdot\|_2$. 特别是, $\|\cdot\| = \|\cdot\|^D$ 当且仅当 $\|\cdot\|$ 是 Euclid 范数 $\|\cdot\|_2$.

证明: 如果 $\|\cdot\| = \sqrt{c}\|\cdot\|_2$, 且 $x \in V$, 则对任意 $x \in V$,

$$\|x\|^D = \max_{\|y\|=1} |x^*y| = \max_{\|y\|_2=1/\sqrt{c}} |x^*y| = \max_{\|y\|_2=1} \left|x^* \frac{y}{\sqrt{c}}\right|$$

$$= \frac{1}{\sqrt{c}} \max_{\|y\|_2=1} |x^*y| = \frac{1}{\sqrt{c}} \|x\|_2^D = \frac{1}{\sqrt{c}} \|x\|_2 = \frac{1}{c} \|x\|.$$

反之，如果 $\|\ \| = c\|\cdot\|^D$ 对某个 $c>0$ 成立，又 $x\in V$，则(5.4.13)给出不等式

$$\|x\|_2^2 = |x^*x| \leqslant \|x\| \|x\|^D = \frac{1}{c}\|x\|^2,$$

所以 $\|x\| \geqslant \sqrt{c}\|x\|_2$. 如果 $x\neq 0$，可以利用这个不等式证明其反向不等式，因为

$$\frac{1}{c}\|x\| = \|x\|^D = \max_{\|y\|=1} |x^*y| = \max_{y\neq 0}\left|x^*\frac{y}{\|y\|}\right|$$

$$= \max_{y\neq 0}\left|x^*\frac{y}{\|y\|_2}\right|\frac{\|y\|_2}{\|y\|} \leqslant \max_{y\neq 0}\left|x^*\frac{y}{\|y\|_2}\right|\frac{1}{\sqrt{c}}$$

$$= x^*\frac{x}{\|x\|_2}\frac{1}{\sqrt{c}} = \|x\|_2\frac{1}{\sqrt{c}}$$

277

这里，用到了对所有 $y\neq 0$ 有 $\|y\|_2/\|y\| \leqslant 1/\sqrt{c}$的事实：Cauchy-Schwarz 不等式保证，当一个 Euclid 单位向量平行于一个固定的非零向量时，这个给定的非零向量与这个单位向量之间的内积的极大值会出现，因此，$\|x\| \leqslant \sqrt{c}\|x\|_2$ 对所有 $x\in V$ 成立. 它与已证的反向不等式一起证明，$\|x\| = \sqrt{c}\|x\|_2$ 对所有 $x\in V$ 成立，当 $c=1$ 时，最后的论断成立，并且说明 Euclid 范数是仅有的等于自己的对偶范数. □

作为最后一个注释，我们要说明一种有用的观念，在这种观念下，一个向量与一个向量范数一样，也有一个对偶.

5.4.17 定义 设 $x\in \mathbf{C}^n$ 是给定的向量，$\|\cdot\|$ 是 $\mathbf{C}^n$ 上给定的范数. 集合

$$\{y\in \mathbf{C}^n: \|y\|^D\|x\| = y^*x = 1\}$$

称为 x 的关于 $\|\cdot\|$ 的对偶. 如果 y 在 x 的关于范数 $\|\cdot\|$ 的对偶中，则向量的有序对 $(x, y)\in \mathbf{C}^n\times\mathbf{C}^n$ 就称为关于 $\|\cdot\|$ 的对偶对.

由推论(5.5.15)可知，如果 $\|\cdot\|$ 是向量范数，则每个向量 $x\in\mathbf{C}^n$ 关于 $\|\cdot\|$ 的对偶是非空的. 它可能由一个点或多个点组成. 例如，如果 $\|\cdot\| = \|\cdot\|_2$，则每个向量 $x\in\mathbf{C}^n$ 的对偶是向量 x 本身. 另外，如果 $\|\cdot\| = \|\cdot\|_\infty$，则 $x=[0, 1]^T$ 的对偶由一个向量组成，但 $x=[1, 1]^T$ 的对偶包含无限多个向量.

习题

1. 注意，(5.4.5)可等价地叙述为

$$C_m(\|\cdot\|_\alpha, \|\cdot\|_\beta) \leqslant \frac{\|x\|_\beta}{\|x\|_\alpha} \leqslant C_M(\|\cdot\|_\alpha, \|\cdot\|_\beta)$$

其中 $C_m(\cdot, \cdot)$和 $C_M(\cdot, \cdot)$表示与(5.4.5)中的相应范数有关的可能达到的最佳常数. 证明 $C_m(\|\cdot\|_\beta, \|\cdot\|_\alpha) = C_M(\|\cdot\|_\alpha, \|\cdot\|_\beta)^{-1}$.

2. 用 $C_m(\|\cdot\|_\alpha, \|\cdot\|_\beta)$和 $C_m(\|\cdot\|_\beta, \|\cdot\|_\gamma)$表示 $C_m(\|\cdot\|_\alpha, \|\cdot\|_\gamma)$，其中，有关的常数不一定是可能达到的最佳值，对 C_M 做同样的表示.

$\alpha \backslash \beta$	1	2	∞
1	1	$\sqrt{n}$	n
2	1	1	$\sqrt{n}$
∞	1	1	1

3. 验证右边附表给出了 L_1，L_2 和 L_∞ 范数之间的最佳界 $C_M(\|\cdot\|_\alpha, \|\cdot\|_\beta)$；即 $\|x\|_\alpha \leqslant C_M\|x\|_\beta$ 对所有 $x\in\mathbf{C}^n$ 与 α，$\beta=1, 2, \infty$ 成立．在每种情形，举出一个非零向量 x 使得 $\|x\|_\alpha = C_M\|x\|_\beta$，说明所给的界是可能达到的最佳界．关于最佳下界 $\|x\|_\alpha \geqslant C_m\|x\|_\beta$ 的表是什么？**提示**：参看习题 1.

[278]

4. 证明，如果实或复向量空间上的两个范数是等价的，则它们的关系可以用像(5.4.5)中那样的两个常数及一个不等式联系起来．**提示**：考虑在 $\|\cdot\|_\beta$ 的单位球 S 上的 $f(x)=1/\|x\|_\alpha$．如果 f 在 S 上无界，则存在序列 $\{x_N\}\subset S$，且适合 $\|x_N\|_\alpha<\dfrac{1}{N}$ 及 $\|x_N\|_\beta\equiv 1$，这与 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 的等价性相矛盾．注意，这与有限维性及紧性没有任何关系．

5. 证明，(5.4.2)中的函数 f_k 有性质：当 $k\to\infty$ 时，对每个 x 有 $f(x)\to 0$；当 k，$j\to\infty$ 时，$\|f_k-f_j\|_1\to 0$；而对每个 $k\geqslant 2$，存在某个 $J>k$ 使得对所有 $j>J$ 都有 $\|f_k-f_j\|_2>k^{1/2}$．因此，一个序列在某种意义下可能对一种范数是(逐点)收敛的 Cauchy 序列，而对另一种范数则不是 Cauchy 序列.

6. 设 V 是完备的实或复向量空间，设 $\{x^{(k)}\}$ 是 V 中给定的序列，$\|\cdot\|$ 是 V 上给定的范数．如果存在 $M\geqslant 0$，使得 $\sum\limits_{k=1}^{n}\|x^{(k)}\|\leqslant M$ 对所有 $n=1, 2, \cdots$ 成立，证明用 $y^{(n)}=\sum\limits_{k=1}^{n}x^{(k)}$ 定义的部分和序列收敛于 V 的一个点．这推广了关于实数的无穷级数的收敛性的什么定理？

7. 证明对每个 $x\in\mathbf{C}^n$ 有 $\|x\|_\infty=\lim\limits_{p\to\infty}\|x\|_p$.

8. 如果 $\alpha>0$ 且 $\|\cdot\|_\alpha=\alpha\|\cdot\|$，证明 $(\|\cdot\|_\alpha)^D=(1/\alpha)\|\cdot\|^D$.

9. 证明，对任意 $p\geqslant 1$，l_p 范数的对偶范数是 l_q 范数，这里，q 是由关系 $1/p+1/q=1$ 定义的．**提示**：用 Hölder 不等式的一般形式代替(5.4.14).

10. 设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 $\mathbf{C}^n$ 上两个给定的范数，并且假定存在某个 $C>0$ 使得 $\|x\|_\alpha\leqslant C\|x\|_\beta$ 对所有 $x\in\mathbf{C}^n$ 成立．证明 $\|x\|_\beta^D\leqslant C\|x\|_\alpha^D$ 对所有 $x\in\mathbf{C}^n$ 成立．**提示**：

$$
\begin{aligned}
\|x\|_\alpha^D &= \max_{\|y\|_\alpha=1}|y^*x| = \max_{y\neq 0}\left|\frac{y^*x}{\|y\|_\alpha}\right| \\
&= \max_{y\neq 0}\frac{|y^*x|}{\|y\|_\alpha} \geqslant \max_{y\neq 0}\frac{|y^*x|}{C\|y\|_\beta} = \frac{1}{C}\max_{\|y\|_\beta=1}|y^*x|.
\end{aligned}
$$

[279]

11. 证明，$\|\cdot\|^D$ 的等距变换群总包含 $\|\cdot\|$ 的所有等距变换的伴随组成的集合．由此推出，$\|\cdot\|^D$ 的等距变换群恰好是由 $\|\cdot\|$ 的等距变换的伴随组成的集合．什么时候 $\|\cdot\|$ 与 $\|\cdot\|^D$ 的等距变换是一样的？

12. 设 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上的向量范数，且设 $T\in M_n$．证明，如果 T 是关于 $\|\cdot\|$ 的等距变换，则 $\|\cdot\|_T^D=\|\cdot\|^D$.

13. 设 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上一个给定的范数．(a)说明零向量关于 $\|\cdot\|$ 的对偶总是 $\{0\}$．(b)试用推论(5.5.15)证明，每个 $x\in\mathbf{C}^n$ 的对偶是非空集合．**提示**：如果 $\|y_0\|^D=1$ 且 $y_0^*x=\|x\|$，确定 $c\geqslant 0$ 使 $y=cy_0$ 在 x 的对偶中．(c)设 $\|\cdot\|$ 是 Euclid 范数 $\|\cdot\|_2$．证明每个 $x\in\mathbf{C}^n$ 的对偶是

$\{x\}$. (d)设 $\|\cdot\| = \|\cdot\|_\infty$. 证明 $x=\begin{bmatrix}0\\1\end{bmatrix}$的对偶是 x; $x=\begin{bmatrix}1\\1\end{bmatrix}$的对偶是从$\begin{bmatrix}0\\2\end{bmatrix}$到$\begin{bmatrix}2\\2\end{bmatrix}$再到$\begin{bmatrix}2\\0\end{bmatrix}$的两条线段. (e)设 $\|\cdot\| = \|\cdot\|_1$. 证明, $x=\begin{bmatrix}2\\2\end{bmatrix}$的对偶是$\left\{\begin{bmatrix}1\\1\end{bmatrix}\right\}$, 而 $x=\begin{bmatrix}0\\1\end{bmatrix}$的对偶是从$-\begin{bmatrix}-1\\1\end{bmatrix}$到$\begin{bmatrix}0\\1\end{bmatrix}$再到$-\begin{bmatrix}1\\1\end{bmatrix}$的两条线段. (f)证明, y 在 x 的关于 $\|\cdot\|$ 的对偶中当且仅当 x 在 y 的关于 $\|\cdot\|^D$ 的对偶中. (g)证明, 对每个 $x\in\mathbf{C}^n$, x 关于 $\|\cdot\|$ 的对偶是$\{x\}$, 当且仅当 $\|\cdot\|_2$.

14. 考虑由 $f(x)=|x_1x_2|^{1/2}$给出的函数 f: $\mathbf{R}^2\to\mathbf{R}$. 证明集合$\{x: f(x)=1\}$是非紧的. 这与(5.4.8)矛盾吗?

15. 考虑正文中给出的 $\mathbf{R}^2$ 上的一个准范数 $f(x)=(\|x\|_\alpha\|x\|_\beta)^{1/2}$的例子, 其中

$$\|x\|_\alpha = \|[10x_1, x_2]^T\|_\infty$$
$$\|x\|_\beta = \|[x_1, 10x_2]^T\|_\infty$$

证明, "单位球"$\{x\in\mathbf{R}^2: f(x)\leqslant 1\}$在第一象限内的部分以线段 $x_2=1/\sqrt{10}$, 线段 $x_1=1/\sqrt{10}$ 和一段双曲线弧 $x_1x_2=\dfrac{1}{100}$为边界. 试画出这个集合的草图并且证明它不是凸集. 为什么该"单位球"在其余三个象限中的余下部分可以通过这个集合依次对各坐标轴的反射来得到? 证明, 对偶范数的单位球$\{x\in\mathbf{R}^2: f^D(x)\leqslant 1\}$在第一象限以线段 $x_1/10+x_2=\sqrt{10}$和线段 $x_1+x_2/10=\sqrt{10}$为边界, 而 f^D 的整个单位球可以通过它在第一象限的部分依次进行反射得到, 并且它是凸集. 证明, f^{DD}的单位球在第一象限内的部分以线段 $x_2=1/\sqrt{10}$, 线段 $x_1=1/\sqrt{10}$和线段 $x_1+x_2=11/(10\sqrt{10})$为边界. 该单位球的余下部分可以通过这个集合依次对各坐标轴的反射来得到, 且它是凸集. 最后把 f^{DD}的单位球与 f 的单位球作比较, 证明前者正好是后者的闭凸包.

16. 设 $\|\cdot\|$ 是 $V=\mathbf{R}^n$ 或 $\mathbf{C}^n$ 上的向量范数. 证明

$$\max_{x\neq 0}\frac{\|x\|^D}{\|x\|} = \max_{\|x\|=1}\max_{\|y\|=1}\left(\frac{x}{\|x\|_2}\right)^*\left(\frac{y}{\|y\|_2}\right)\|x\|_2\|y\|_2$$
$$\leqslant\left[\max_{\|x\|=1}\|x\|_2\right]^2\equiv C_M$$

且

$$\min_{x\neq 0}\frac{\|x\|^D}{\|x\|}\geqslant\left[\min_{\|x\|=1}\|x\|_2\right]^2\equiv C_m$$

证明 $C_m\|x\|\leqslant\|x\|^D\leqslant C_M\|x\|$ 对所有 $x\in V$ 成立. 所以, 几何常数给出每个范数与其对偶范数之间的这两个界.

17. 设 $f(\cdot)$是 $\mathbf{R}^n$ 或 $\mathbf{C}^n$ 上的准范数, 证明

$$f^D(y)=\max_{f(x)=1}\mathrm{Re}\,y^*x=\max_{f(x)=1}|y^*x|$$
$$=\max_{x\neq 0}\frac{\mathrm{Re}\,y^*x}{f(x)}=\max_{x\neq 0}\frac{|y^*x|}{f(x)} \tag{5.4.18}$$

关于这种想法的另一个例子可参看(5.6.1)下面的练习.

进一步阅读 关于各种对偶范数的讨论见[Hou 64]. 一个准范数的对偶是一个范数的思想似乎应属于 J. Von Neumann, 他在下述文章中讨论了度规函数(我们称之为向量范数): "Some Matrix-Inequalities and Metrization of Matric-Space," *Tomsk Univ. Rev.* 1(1937),

280

205-218. 这篇文章更容易从下述 Von Neumann 的著作中找到：*Collected Works*, *vol.* 4, ed. A. H. Taub, Macmillian, New York, 1962.

5.5 向量范数的几何性质

向量范数的原始几何特征是它的单位球，通过它，可以透彻理解有关范数的重要性质.

5.5.1 定义 设 $\|\cdot\|$ 是实或复向量空间 V 上的向量范数，x 是 V 的一个点，且设 $r>0$ 是给定的. 以 x 为中心，r 为半径的球是集合

$$B_{\|\cdot\|}(r;x)\equiv\{y\in V:\|y-x\|\leqslant r\}.$$

$\|\cdot\|$ 的单位球是集合

$$B_{\|\cdot\|}\equiv B_{\|\cdot\|}(1;0)=\{y\in V:\|y\|\leqslant 1\}.$$

[281] **练习** 证明，对每个 $r>0$ 以及每个 $x\in V$，$B(r;x)=\{x+y:y\in B(r;0)\}=x+B(r;0)$.

一个以任意点 x 为中心的有给定半径的球与以零点为中心的有相同半径的球看做是一样的；这只要把零点平移到 x 点就可以了. 单位球是范数的一个几何缩影，因为齐次性，单位球刻划了范数的特征(实际上只需要 $B_{\|\cdot\|}$ 的边界). 现在我们要确定，究竟 $\mathbf{C}^n$ 的哪些子集可以是某个向量范数的单位球.

练习 画出 $\mathbf{R}^2$ 上的 l_1，l_2 和 l_∞ 范数的单位球的草图. 它们之间是否存在包含关系？哪些点一定在 $\mathbf{R}^2$ 上的任一 l_p 范数的单位球的边界上？画出几个 l_p 范数的单位球的草图.

练习 如果 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是向量空间 V 上的两个范数，证明，$\|x\|_\alpha\leqslant\|x\|_\beta$ 对所有 $x\in V$ 成立，当且仅当 $B_{\|\cdot\|_\beta}\subset B_{\|\cdot\|_\alpha}$. 因此向量范数的自然偏序关系可以用几何的包含关系来表示. 当一个范数乘以一个正常数时，单位球会发生什么变化？

练习 如果 $\|\cdot\|$ 是 V 上的向量范数，又如果 $x\in V$，且 α 是使 $\|\alpha x\|=\|x\|$ 的纯量，证明 $\alpha=0$ 或 $|\alpha|=1$，由此得出，每条"射线"$\{\alpha x:\alpha>0\}$ 与 $\|\cdot\|$ 的单位球的边界恰好相交一次.

5.5.2 定义 一个范数称为是多面的，是指它的单位球是多面体.

练习 哪些 l_p 范数是多面的？

练习 如果 $\|\cdot\|$ 是多面范数，如果 $S\in M_n$ 是非奇异矩阵，$\|\cdot\|_S$ 是多面范数吗？

开集和闭集这些基本的拓扑概念在具有范数的向量空间中是很容易定义的.

5.5.3 定义 设 $\|\cdot\|$ 是实或复向量空间 V 上的范数，S 是 V 的子集，点 $x\in S$ 称为 S 的内点，是指存在 $\varepsilon>0$ 使 $B(\varepsilon;x)\subset S$. 集合 S 称为开集，是指 S 的每点都是 S 的内点；S 称为闭集，是指它的补集是开集. S 的极限点是这样一个点 $x\in V$，它对某个序列 $x^{(k)}\subset S$，(关于 $\|\cdot\|$)有 $\lim\limits_{k\to\infty}x^{(k)}=x$. S 的闭包是 S 与它的极限点集合的并集. S 的边界是 S 的闭包与 S 的补集的闭包的交. 集合 S 是有界的，指的是，存在某个 $M>0$，使得 $S\subset B_{\|\cdot\|}(M;0)$，集合 S 是紧的，如 [282] 果可以从每个用开集作成的复盖 $\bigcup_\alpha S_\alpha\supset S$ 中选出有限多个集合 $S_{\alpha_1},\cdots,S_{\alpha_N}$ 使得 $\bigcup_{i=1}^N S_{\alpha_i}\supset S$.

练习 证明任意实或复向量空间 V 上的任何范数的单位球 $B_{\|\cdot\|}$ 是有界闭集.

练习 设 V 是有限维实或复向量空间，$S\subset V$ 是有界闭集. 利用对某个 n 有 V 同构于 $\mathbf{R}^n$ 或 $\mathbf{C}^n$ 的事实(见附录 E)证明 S 是紧集.

5.5.4 论断 如果$\|\cdot\|$是非平凡(即非零维)的实或复向量空间V上的向量范数，则0是单位球$B_{\|\cdot\|}$的内点. 这可由范数$\|\cdot\|$的齐次性及正定性得到，它蕴涵$B_{\|\cdot\|}\left(\frac{1}{2};0\right)\subset B_{\|\cdot\|}(1;0)$；且前者的边界处在后者的内部.

5.5.5 论断 向量范数的单位球是均衡的；也就是说，如果x在这个单位球中，则对适合$|\alpha|=1$的所有纯量α，αx也在该单位球中. 这一结论可由向量范数的齐次性得到.

5.5.6 论断 有限维向量空间上的向量范数的单位球是紧集，因为向量范数的齐次性它是有界的；因为范数总是连续函数，它又是闭的. 在有限维情形，有界闭集是紧集，但在无维情形，有界闭集不总是紧的. 要经常用到的紧集性质是 Weierstrass 定理(附录 E)：紧集上的连续实值函数是有界的，并且在该集上达到它的上确界和下确界. 因为这个理由，我们经常提及这种函数的“极大值”和“极小值”.

练习 考虑由具有可数多个分量的向量$x=(x_i)$组成的复向量空间l_2，它的范数是有限范数的自然推广

$$\|x\|_2=\left(\sum_{k=1}^{\infty}|x_k|^2\right)^{1/2}.$$

证明，对每对不同的单位基向量e_k和e_j，j，$k=1$，2，…，有$\|e_k-e_j\|_2=\sqrt{2}$. 因此$\{e_k\}$的任何无穷子序列不可能是 Cauchy 序列，所以不可能有任何收敛的子序列. 由此可知l_2的单位球不可能是紧集.

283

5.5.7 论断 向量范数的单位球是凸集.

证明：如果$\|x\|\leqslant 1$，$\|y\|\leqslant 1$，且$\alpha\in[0,1]$，则

$$\|\alpha x+(1-\alpha)y\|\leqslant\|\alpha x\|+\|(1-\alpha)y\|=\alpha\|x\|+(1-\alpha)\|y\|\leqslant\alpha+(1-\alpha)\leqslant 1,$$

所以$\alpha x+(1-\alpha)y$也位于单位球中. □

上述关于范数的单位球的这些必要条件也是刻划范数特征的充分条件.

5.5.8 定理 有限维实或复向量空间中的集合B是V上一个向量范数的单位球，当且仅当B是：(i)紧集，(ii)凸集，(iii)均衡集，(iv)以O作为内点.

证明：我们已经得知条件(i)～(iv)是必要的，为了看出它们对范数定义是充分的，考虑任意非零点$x\in V$，作从原点到x的一条射线线段$\{\alpha x: 0<\alpha\leqslant 1\}$，在这条射线上，用原点到单位球的边界上的唯一一点间的线段长度作为一个单位，于是x的“长度”定义为沿该射线从原点到x的比例距离，更形式地，定义$\|x\|$为：

$$\begin{cases}\|x\|=0, & \text{如果 } x=0,\\ \|x\|=\min\left\{\dfrac{1}{t}: t>0 \text{ 且 } tx\in B\right\}, & \text{如果 } x\neq 0.\end{cases}$$

这个函数是有意义的，有限的，且对每个非零向量x是正的，这是因为B是紧集，且0是B的内点. 利用均衡性假设容易看出$\|\cdot\|$是齐次函数，所以，余下只需验证三角不等式. 如果x与y是给定的非零向量，则$x/\|x\|$与$y/\|y\|$是B的边界上的单位向量. 由凸性，向量

$$z=\frac{\|x\|}{\|x\|+\|y\|}\frac{x}{\|x\|}+\frac{\|y\|}{\|x\|+\|y\|}\frac{y}{\|y\|}$$

也一定在 B 中. 因此, $\|z\|\leqslant 1$, 且易算出, 这等价于 $\|x+y\|\leqslant\|x\|+\|y\|$. □

练习 给出(5.5.8)的证明细节, 仔细注意四个假设中的每一个各用在何处.

我们熟悉的所有 l_p 向量范数都有这样的性质, 那就是 $\|x\|$ 只依赖于 x 的各元的绝对值, 此外, 每个 l_p 范数是 x 的各元的绝对值的递增函数. 这两个性质不是没有关系的. [284]

5.5.9 定义 如果 $x=[x_i]\in\mathbf{F}^n(\mathbf{R}^n$ 或 $\mathbf{C}^n)$, 定义 $|x|\equiv[|x_i|]$. 我们说 $|x|\leqslant|y|$, 是指 $|x_i|\leqslant|y_i|$ 对所有 $i=1,2,\cdots,n$ 成立. $\mathbf{F}^n$ 上的向量范数 $\|\cdot\|$ 称为

(a) 单调的, 指的是对所有 $x,y\in\mathbf{F}^n$, $|x|\leqslant|y|$ 蕴涵 $\|x\|\leqslant\|y\|$;

(b) 绝对的, 指的是对所有 $x\in\mathbf{F}^n$ 有 $\|x\|=\||x|\|$.

5.5.10 定理 $\mathbf{F}^n(\mathbf{R}^n$ 或 $\mathbf{C}^n)$上的范数 $\|\cdot\|$ 是单调的当且仅当它是绝对的.

证明: 如果 $\|\cdot\|$ 是单调的, 且 $x\in\mathbf{F}^n$, 设 $y\equiv|x|$. 于是 $|y|\leqslant|x|$, 且 $|x|\leqslant|y|$, 所以 $\|y\|\leqslant\|x\|$ 且 $\|x\|\leqslant\|y\|$, 因此 $\|\cdot\|$ 是绝对的. 如果 $\|\cdot\|$ 是绝对的, 设 $x=[x_i]\in\mathbf{F}^n$ 是给定的向量, 设 k 是给定的整数, 且 $1\leqslant k\leqslant n$, 又设 $\alpha\in[0,1]$. 则

$$
\begin{aligned}
&\|[x_1,\cdots,x_{k-1},\alpha x_k,x_{k+1},\cdots,x_n]^T\|\\
&=\left\|\frac{1}{2}(1-\alpha)[x_1,\cdots,x_{k-1},-x_k,x_{k+1},\cdots,x_n]^T+\frac{1}{2}(1-\alpha)x+\alpha x\right\|\\
&\leqslant\frac{1}{2}(1-\alpha)\|[x_1,\cdots,x_{k-1},-x_k,x_{k+1},\cdots,x_n]^T\|+\frac{1}{2}(1-\alpha)\|x\|+\alpha\|x\|\\
&=\frac{1}{2}(1-\alpha)\|x\|+\frac{1}{2}(1-\alpha)\|x\|+\alpha\|x\|=\|x\|.
\end{aligned}\tag{5.5.11}
$$

范数是绝对的这一假设条件只用在倒数第二个等式中. 对各不相同的分量重复应用(5.5.11), 可以证明绝对范数有性质: 对每个 $x\in\mathbf{F}^n$ 及 $\alpha_k\in[0,1]$的所有选择, $k=1,\cdots,n$,

$$\|[\alpha_1x_1,\cdots,\alpha_nx_n]^T\|\leqslant\|[x_1,\cdots,x_n]^T\|.\tag{5.5.12}$$

最后, 如果 $|x|\leqslant|y|$, 则对每个 $k=1,2,\cdots,n$, 存在实数 α_k 和 θ_k, 且 $\alpha_k\in[0,1]$, 使 $x_k=\alpha_ke^{i\theta_k}y_k$. 于是, 利用绝对性便有

$$
\begin{aligned}
\|x\|&=\|[\alpha_1e^{i\theta_1}y_1,\cdots,\alpha_ne^{i\theta_n}y_n]^T\|=\|[\alpha_1|y_1|,\cdots,\alpha_n|y_n|]^T\|\\
&\leqslant\|[|y_1|,\cdots,|y_n|]^T\|=\|y\|,
\end{aligned}
$$

所以这个范数一定是单调的. □

不等式(5.5.11)启发我们提出稍弱的单调性概念.

[285] **5.5.13 定义** $\mathbf{F}^n(\mathbf{R}^n$ 或 $\mathbf{C}^n)$上的向量范数 $\|\cdot\|$ 称为弱单调的, 是指

$$\|[x_1,\cdots,x_{k-1},0,x_{k+1},\cdots,x_n]^T\|\leqslant\|[x_1,\cdots,x_{k-1},x_k,x_{k+1},\cdots,x_n]^T\|$$

对所有 $x\in\mathbf{F}^n$ 及所有 $k=1,2,\cdots,n$ 成立.

如果范数 $\|\cdot\|$ 是弱单调的, 又如果 $\alpha\in[0,1]$, 则

$$
\begin{aligned}
&\|[x_1,\cdots,x_{k-1},\alpha x_k,x_{k+1},\cdots,x_n]^T\|\\
&=\|(1-\alpha)[x_1,\cdots,x_{k-1},0,x_{k+1},\cdots,x_n]^T+\alpha x\|\\
&\leqslant(1-\alpha)\|[x_1,\cdots,x_{k-1},0,x_{k+1},\cdots,x_n]^T\|+\alpha\|x\|\\
&\leqslant(1-\alpha)\|x\|+\alpha\|x\|=\|x\|,
\end{aligned}
$$

所以弱单调范数满足较强的条件(5.5.12). 因此，如果在弱单调范数的单位球面上给定一点，且该点的一个坐标变到零，则这样产生的整个线段一定在单位球内. 单调范数显然是弱单调的，但是，反之不成立，这正是下面的练习要证明的.

练习 证明顶点在$\pm[2,2]^T$与$\pm[1,-1]^T$的平行四边形是$\mathbf{R}^2$上的一个向量范数的单位球，且这个范数不是弱单调的.

练习 函数$f(x)=|x_1-x_2|+|x_2|$是$\mathbf{R}^2$上的向量范数吗？它是单调的吗？它是弱单调的吗？画出它的单位球的草图.

练习 设$\|\cdot\|$是$\mathbf{R}^2$上的绝对范数，证明，如果$x=[x_1, x_2]^T$是单位球边界上一点，则$[\pm x_1, \pm x_2]^T$(所有四种可能的选择)都在边界上. 举例并画出$\mathbf{R}^2$上的一个非绝对向量范数的单位球的草图，说明它的几何性质. 在$\mathbf{R}^n$中会出现什么情形？

练习 画出$\mathbf{R}^2$中顶点在$\pm[0,1]^T$，$\pm[1,0]^T$和$\pm[1,1]^T$的多边形的草图. 说明它为什么是$\mathbf{R}^n$上的一个弱单调向量范数的单位球，而不是单调或绝对向量范数的单位球.

向量范数的单位球的凸性具有许多往往是令人惊叹的深刻结论. 其中之一是下面的对偶性定理，我们一般从准范数的角度来叙述它. 所涉及的基本思想是很自然的几何思想，也就是说，包含某个集合S的最小闭凸集(闭凸包CoS；见附录B)是包含S的所有闭半空间(在超平面一边的所有点)的交，并且，如果存在这样一点x，只要S位于一个半空间内，x也就位于这同一个半空间内，则x必定属于S的闭凸包. 这些简单的概念直接导出一个重要结论，一个向量范数的两次对偶等于原范数. 286

5.5.14 定理(对偶性定理) 设$f(\cdot)$是$V=\mathbf{R}^n$或$\mathbf{C}^n$上的准范数，设f^D表示f的对偶范数，而f^{DD}是f^D的对偶范数，又设

$$B\equiv\{x\in V: f(x)\leqslant 1\},$$
$$B''\equiv\{x\in V: f^{DD}(x)\leqslant 1\}$$

分别表示f的"单位球"和f^{DD}的单位球. 则

$$B\subset B''=\text{Co}\,B,$$

因而$f^{DD}(x)\leqslant f(x)$对所有$x\in V$成立. 如果f是V上的向量范数，则$B=B''$且$f^{DD}=f$.

证明：设$x\in V$是给定的向量，则(5.4.13)说明，对任意$y\in V$，

$$|y^*x|\leqslant f(x)f^D(y),$$

因而

$$f^{DD}(x)=\max_{f^D(y)=1}|y^*x|\leqslant\max_{f^D(y)=1}f(x)f^D(y)=f(x).$$

因此，$f^{DD}(x)\leqslant f(x)$对所有$x\in V$成立，这个不等式等价于几何命题$B\subset B''$.

为了证明第二个包含关系，采用对偶范数的特征(5.4.18)是方便的，还应知道集合$\{t\in V:\text{Re}\,t^*v\leqslant 1\}$是包含原点一般闭半空间. 利用对偶的范数的定义，设$u\in B''$是给定的点，我们看出，

$$\begin{aligned}u\in&\{t:\text{Re}\,t^*v\leqslant 1,\text{对每个适合 } f^D(v)\leqslant 1 \text{ 的 } U\}\\&=\{t:\text{Re}\,t^*v\leqslant 1,\text{对每个适合 } f(w)\leqslant 1 \text{ 的 } w \text{ 的每个适合 } \text{Re}\,v^*w\leqslant 1 \text{ 的 } v\}\\&=\{t:\text{Re}\,t^*v\leqslant 1,\text{对所有 } w\in B \text{ 的每个适合 } \text{Re}\,w^*v\leqslant 1 \text{ 的 } v\}\end{aligned}$$

这说明 u 位于包含 B 的各个点的每个闭半空间中：即 u 位于每个包含 B 的闭半空间中. 因为所有这些闭半空间的交是 B 的闭凸包 $\mathrm{Co}\,B$，得知 $u\in \mathrm{Co}\,B$. 然而点 $u\in B''$是任意的，所以 $B''\subset \mathrm{Co}\,B$，因为 $\mathrm{Co}\,B$ 是包含 B 的所有凸集的交，我们也有 $\mathrm{Co}\,B\subset B''$，因而 $B''=\mathrm{Co}\,B$.

如果准范数 f 实际上是范数，则它的闭单位球 B 是凸集，因而 $B=\mathrm{Co}\,B$，所以 $B\subset B''\subset$ 287 B；因而 $B=B''$. 因为它们的单位球是恒等的，所以范数 f 与 f^{DD} 相同. □

对偶性定理的一个应用是下述有用的结果. 它是泛函分析中的一个重要的一般结果(称之为 Hahn-Banach 定理)关于有限维形式的特殊情形.

5.5.15 推论 设 $y\in \mathbf{C}^n$ 是给定的向量，$\|\cdot\|$ 是 $\mathbf{C}^n$ 上给定的向量范数，则存在向量 $y_0\in \mathbf{C}^n$ 使得

(a) $|(y_0)^*x|\leqslant \|x\|$ 对所有 $x\in \mathbf{C}^n$ 成立；

(b) $(y_0)^*y=\|y\|$.

向量 y_0 不一定唯一，不过 $\|y_0\|^D=1$ 且$(y_0)^*y=\|y\|$.

证明： 我们知道，根据对偶性定理，

$$\|y\| = (\|y\|^D)^D = \max_{\|z\|^D=1} |y^*z|,$$

并且由向量范数 $\|\cdot\|^D$ 的单位球面的紧性可知，该极大值实际上可由某个(不一定唯一的)适合 $\|y_0\|^D=1$ 的 $z=y_0$ 达到，因而 $\|y\|=|y^*y_0|$. 用模为 1 的适当因子乘 y_0 便知，可以使内积 y^*y_0 为正值，因而(b)已被证明. 一般，从(5.4.13)可知，对所有 $x\in \mathbf{C}^n$ 有

$$|(y_0)^*x|\leqslant \|y_0\|^D\|x\| = \|x\|.$$

因此向量 y_0 同样适合(a). 注意，(a)说明 $\|y_0\|^D\leqslant 1$，而(b)使 $\|y_0\|^D=1$. □

习题

1. 证明，集合 S 是闭的，当且仅当它包含它的所有极限点.

2. 证明，S 的每一点是 S 的极限点，因而 S 的闭包恰好是由 S 的极限点组成的集合.

3. 给出一个既是开集又是闭集的集合例子. 给出一个既不是开集又不是闭集的集合例子.

4. 设 S 是具有范数 $\|\cdot\|$ 的实或复向量空间 V 中的紧集. 证明 S 是有界闭集. 如果$\{x_\alpha\}$是给定的无穷序列，证明，存在一个可数子序列$\{x_{\alpha_i}\}\subset\{x_\alpha\}$和一点 $x\in S$，使得$\lim_{i\to\infty} x_{\alpha_i}=x$. 证明紧集的任意闭子集是紧集.

5. 在(5.5.4)中，如果 V 是零维的，会出现什么情形？

6. 如何定义向量半范数的单位球，它的形状与范数的单位球有何不同？画出一个例子的 288 草图.

7. 如果 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是一个向量空间上的向量范数，又如果 $\|\cdot\|$ 是用

$$\|x\| = \max\{\|x\|_\alpha, \|x\|_\beta\}$$

定义的向量范数，证明 $B_{\|\cdot\|}=B_{\|\cdot\|_\alpha}\cap B_{\|\cdot\|_\beta}$.

8. 证明，$\mathbf{F}^n$($\mathbf{R}^n$ 或 $\mathbf{C}^n$)上的向量范数 $\|\cdot\|$ 是绝对的，当且仅当

$$\|[\alpha_1x_1,\alpha_2x_2,\cdots,\alpha_nx_n]^T\| = \|[x_1,x_2,\cdots,x_n]^T\|$$

对所有$[x_1, x_2, \cdots, x_n]^T\in \mathbf{F}^n$ 和所有适合 $|\alpha_1|=\cdots=|\alpha_n|=1$ 的纯量 $\alpha_1, \alpha_2, \cdots, \alpha_n\in \mathbf{F}$

成立.

在以下六个题中，要用到下述记号. 设 x，$y\in V$，设 $\|\cdot\|$ 是实或复向量空间 V 上的向量范数. 则

$$L(x,y)\equiv\{z(t)=x+t(y-x):0\leqslant t\leqslant 1\}$$

表示 x 与 y 间的普通(线性代数)线段，且

$$C(x,y;\|\cdot\|)\equiv\{z\in V:\|x-z\|+\|z-y\|=\|x-y\|\}$$

表示关于范数 $\|\cdot\|$ 的 x 和 y 的(度量)凸包.

9. 证明 $L(x,y)\subset C(x,y;\|\cdot\|)$ 对所有 x，$y\in V$ 和任何向量范数 $\|\cdot\|$ 成立.

10. 如果 $V=\mathbf{C}^n$，且其范数是 l_2 范数，证明 $C(x,y;\|\cdot\|_2)=L(x,y)$ 对所有 x，$y\in\mathbf{C}^n$ 成立；即证明，$\|x+y\|_2=\|x-z\|_2+\|z-y\|_2$，当且仅当 $z=x+t(y-x)$ 对某个 $t\in[0,1]$ 成立.

11. 证明 $C(x,y;\|\cdot\|)$ 总是一个普通凸集；即证明，如果 z_1，$z_2\in C(x,y;\|\cdot\|)$，则 $tz_1+(1-t)z_2\in C(x,y;\|\cdot\|)$ 对所有 $t\in[0,1)$ 成立.

12. 考虑 $\mathbf{R}$ 上的 $V=\mathbf{R}^2$，证明 $C((1,0),(0,1);\|\cdot\|_1)$ 是平面中顶点在点 $(0,0)$，$(0,1)$，$(1,1)$ 和 $(1,0)$ 的整个正方形. **提示**：证明 $(0,0)$ 和 $(1,1)$ 在 $(0,1)$ 和 $(1,0)$ 的 l_1 凸包中，然后利用习题 11. 另外证明 $C((1,0),(0,1);\|\cdot\|_\infty)$ 恰好是线段 $L((1,0),(0,1))$.

13. 再考虑 $\mathbf{R}$ 上的 $V=\mathbf{R}^2$，证明 $C((1,1),(1,-1);\|\cdot\|_\infty)$ 是平面中顶点 $(0,0)$，$(1,1)$，$(2,0)$ 和 $(1,-1)$ 的整个正方形. **提示**：证明 $(0,0)$ 和 $(2,0)$ 在 $(1,1)$ 和 $(1,-1)$ 的 l_∞ 凸包中. 另外证明 $C((1,1),(1,-1);\|\cdot\|_1)$ 恰好是线段 $L((1,1),(1,-1))$.

14. 有 $k(\geqslant 2)$ 个点的点组 $S\subset V$ 的度量凸包可以定义为所有这样一些 $z\in V$ 的点组成的集合，使得 z 在两个点的度量凸包中，而这两个点中的每一点又在 S 的某一对点的度量凸包中. 证明，当 $k=2$ 时，这与上述定义是一致的，描述 $\mathbf{R}^n$ 中的一组标准正交基向量 $\{e_1,e_2,\cdots,e_n\}$ 的 l_1 凸包. 这个基的 l_2 凸包是什么？这个基的普通线性代数凸包是什么？ [289]

进一步阅读 关于向量范数的几何特性的进一步讨论见[Hou 64]. Von Neumann 在(5.4)节末引用的文章中采用了证明对偶性定理的关键思想(一个范数或准范数的单位球等同于所有包含该范数或准范数的单位球的所有半空间之交). 关于凸集，凸包，半空间等等的详细讨论见[Val].

5.6 矩阵范数

因为 M_n 本身是 n^2 维向量空间，所以可以采用 $\mathbf{C}^{n^2}$ 上的任一种向量范数来度量矩阵的“大小”. 但是，M_n 不仅仅是高维向量空间；它还有通常的乘法运算，并且在度量矩阵的“大小”时，建立起 AB 的“大小”与 A 和 B 的“大小”间的关系常常是有用的.

我们称函数 $\|\cdot\|$：$M_n\rightarrow\mathbf{R}$ 是矩阵范数，指的是，对所有 A，$B\in M_n$，它满足下列五条公理：

(1) $\|A\|\geqslant 0$，　　　　非负性；

(1a) $\|A\|=0$，当且仅当 $A=0$，　　　　正定性；

(2) $\|cA\|=|c|\,\|A\|$ 对所有复纯量 c 成立，　　　　齐次性；

(3) $\|A+B\|\leqslant\|A\|+\|B\|$，　　　　三角不等式；

(4) $\|AB\|\leqslant\|A\|\,\|B\|$，　　　　次乘性.

注意，性质(1)～(3)与向量范数(5.1.1)的公理相同. 关于矩阵的向量范数，即满足(1)～(3)而不一定满足(4)的函数，称为广义矩阵范数. 矩阵半范数和广义矩阵半范数的概念也可以通过取消公理(1a)来定义.

对任意矩阵范数有 $\|A^2\|=\|AA\|\leqslant\|A\|\,\|A\|=\|A\|^2$，所以对适合 $A^2=A$ 的任意非零矩阵 A，一定有 $\|A\|\geqslant 1$. 特别是 $\|I\|\geqslant 1$ 对任何矩阵范数成立. 如果 A 是可逆矩阵，则 $I=AA^{-1}$，因而，$\|I\|=\|AA^{-1}\|\leqslant\|A\|\,\|A^{-1}\|$，且有下界

$$\|A^{-1}\|\geqslant\frac{\|I\|}{\|A\|},$$

290 它对任何矩阵范数 $\|\cdot\|$ 都成立.

练习　证明，如果 $\|\cdot\|$ 是矩阵范数，则 $\|A^k\|\geqslant\|A\|^k$ 对每个 $k=1, 2, \cdots$ 和所有 $A\in M_n$ 成立. 给出一个例子，使得这个不等式对关于矩阵的向量范数不成立.

当把在(5.2)中引进的某些向量范数应用于向量空间 M_n 时，它们就是矩阵范数，而有些向量范数则不是矩阵范数. 最熟悉的例子是 $p=1, 2, \infty$ 时的 l_p 范数，已经知道它们是向量范数，所以只需要验证公理(4).

例　对 $A\in M_n$，用

$$\|A\|_1\equiv\sum_{i,j=1}^{n}|a_{ij}|$$

定义的 l_1 范数是矩阵范数，这是因为

$$\begin{aligned}\|AB\|_1&=\sum_{i,j=1}^{n}\left|\sum_{k=1}^{n}a_{ik}b_{kj}\right|\leqslant\sum_{i,j,k=1}^{n}|a_{ik}b_{kj}|\\&\leqslant\sum_{i,j,k,m=1}^{n}|a_{ik}b_{mj}|=\left(\sum_{i,k=1}^{n}|a_{ik}|\right)\left(\sum_{j,m=1}^{n}|b_{mj}|\right)\\&=\|A\|_1\,\|B\|_1.\end{aligned}$$

第一个不等式可以从三角不等式得到，而第二个不等式是把一些附加项加到和中得来的.

例　对 $A\in M_n$，用

$$\|A\|_2\equiv\left(\sum_{i,j=1}^{n}|a_{ij}|^2\right)^{1/2}$$

定义的 Euclid 范数或 l_2 范数是矩阵范数，这是因为

$$\|AB\|_2^2=\sum_{i,j=1}^{n}\left|\sum_{k=1}^{n}a_{ik}b_{kj}\right|^2\leqslant\sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}|^2\right)\left(\sum_{m=1}^{n}|b_{mj}|^2\right)$$

$$= \left(\sum_{i,k=1}^{n} |a_{ik}|^2\right)\left(\sum_{m,j=1}^{n} |b_{mj}|^2\right) = \|A\|_2^2 \|B\|_2^2.$$

这个不等式正是 Cauchy-Schwarz 不等式．当应用于矩阵时，这个范数有时称为 Frobenius 范数，Schur 范数或 Hilbert-Schmidt 范数．如果用 $A \in M_n$ 的列向量 $a_i \in \mathbf{C}^n$ 表示 $A = [a_1 a_2 \cdots a_n] \in M_n$，则

$$\|A\|_2^2 = \|a_1\|_2^2 + \cdots + \|a_n\|_2^2.$$

因为 $\mathbf{C}^n$ 上的 l_2 范数是酉不变的，我们有重要的事实： [291]

$$\|UA\|_2^2 = \|Ua_1\|_2^2 + \cdots + \|Ua_n\|_2^2 = \|a_1\|_2^2 + \cdots + \|a_n\|_2^2 = \|A\|_2^2,$$

其中 $U \in M_n$ 是任意酉矩阵．因为 $\|B^*\|_2 = \|B\|_2$ 对所有 $B \in M_n$ 成立，这蕴涵

$$\|UAV\|_2 = \|AV\|_2 = \|V^* A^*\|_2 = \|A^*\|_2 = \|A\|_2,$$

其中 $U, V \in M_n$ 是任意酉矩阵．因此，M_n 上的 l_2 范数是酉不变矩阵范数．

例 对 $A \in M_n$，用

$$\|A\|_\infty \equiv \max_{1 \le i,j \le n} |a_{ij}|$$

定义的 l_∞ 范数是向量空间 M_n 上的范数，但不是矩阵范数．考虑矩阵 $J = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \in M_2$，由计算可知，$J^2 = 2J$，$\|J\|_\infty = 1$，$\|J^2\|_\infty = \|2J\|_\infty = 2\|J\|_\infty = 2$．它不适合 $\|J^2\|_\infty \le \|J\|_\infty^2$，因而 $\|\cdot\|_\infty$ 不是次乘性范数．但是，如果定义

$$|||A||| = n\|A\|_\infty, A \in M_n,$$

则有

$$|||AB||| = n \max_{1 \le i,j \le n} \left|\sum_{k=1}^{n} a_{ik} b_{kj}\right| \le n \max_{1 \le i,j \le n} \sum_{k=1}^{n} |a_{ik} b_{kj}|$$

$$\le n \max_{1 \le i,j \le n} \sum_{k=1}^{n} \|A\|_\infty \|B\|_\infty = n\|A\|_\infty n\|B\|_\infty$$

$$= |||A||| \, |||B|||.$$

因此，只需要对向量范数 $\|\cdot\|_\infty$ 作稍许改动就可以使它成为矩阵范数．

与 $\mathbf{C}^n$ 上的每个向量范数 $\|\cdot\|$ 相关联的矩阵范数 $|||\cdot|||$ 自然是由 M_n 上的 $\|\cdot\|$ "诱导"的矩阵范数．该范数 $|||\cdot|||$ 是由 $\|\cdot\|$ 构造出来的，而这种构造法也是从一种范数产生另一种范数的方法．

5.6.1 定义 设 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上的向量范数．定义 M_n 上的 $|||\cdot|||$ 为

$$|||A||| \equiv \max_{\|x\|=1} \|Ax\|.$$

上述定义中取"max"（而不是"sup"）是合理的，这是因为 $\|Ax\|$ 是 x 的连续函数且单位球 $B_{\|\cdot\|}$ 是紧集（见附录 E）． [292]

练习 证明范数(5.6.1)也可以用下述等价方式来计算：

$$|||A||| = \max_{\|x\| \le 1} \|Ax\| = \max_{\|x\| = 1} \|Ax\|$$

$$= \max_{x \neq 0} \frac{\| Ax \|}{\| x \|} = \max_{\| x \|_\alpha = 1} \frac{\| Ax \|}{\| x \|},\text{其中 } \| \cdot \|_\alpha \text{ 是任意向量范数}.$$

5.6.2 定理 (5.6.1)中定义的函数$\|\cdot\|$是M_n上的矩阵范数，$\| Ax \| \leqslant \| A \| \| x \|$对所有$A \in M_n$和所有$x \in \mathbf{C}^n$成立，且$\| I \| = 1$.

证明 因为$\| A \|$是非负值函数的极大值，所以本节开头所述公理(1)成立，又$Ax = 0$对所有x成立仅当$A = 0$时才成立，所以公理(1a)成立. 由计算可知，

$$\| cA \| = \max \| cAx \| = \max | c | \| Ax \| = | c | \max \| Ax \| = | c | \| A \|,$$

由此推出公理(2)成立. 类似地，三角不等式(3)被继承下来，这是因为

$$\| A + B \| = \max \| (A + B)x \| = \max \| Ax + Bx \| \leqslant \max(\| Ax \| + \| Bx \|)$$

$$\leqslant \max \| Ax \| + \max \| Bx \| = \| A \| + \| B \|.$$

次乘性公理(4)可从下述事实得出：

$$\| AB \| = \max \frac{\| ABx \|}{\| x \|} = \max \frac{\| ABx \|}{\| Bx \|} \frac{\| Bx \|}{\| x \|}$$

$$\leqslant \max \frac{\| Ay \|}{\| y \|} \max \frac{\| Bx \|}{\| x \|} = \| A \| \| B \|,$$

这里，不妨假定极大值只取遍那些不在B的零空间的x. 对于下一个论断，因为范数$\|\cdot\|$是以取极大值的方式来定义的，由此得知，如果$x \neq 0$，则$\| Ax / \| x \| \| \leqslant \| A \|$. 由向量范数的齐次性可知，$\| Ax \| \leqslant \| A \| \| x \|$，当$x = 0$时，不等式也成立. 最后

$$\| I \| = \max_{\| x \| = 1} \| Ix \| = \max_{\| x \| = 1} \| x \| = 1. \qquad \square$$

293

5.6.3 定义 我们称(5.6.1)中定义的矩阵范数$\|\cdot\|$是由向量范数$\|\cdot\|$诱导的矩阵范数. 有时称它为算子范数或与向量范数$\|\cdot\|$相关联的lub(最小上界)范数.

注意，算子范数是矩阵范数可以作为所有向量范数的一般性质的推论. 因此，一种证明M_n上的某个函数是矩阵范数的方法是，证明它是由某个向量范数所诱导的. 当讨论称之为谱范数的重要矩阵范数时，将采用这种方法.

定理(5.6.2)中所述不等式说明，向量范数$\|\cdot\|$与所诱导的矩阵范数$\|\cdot\|$是相容的，因而这个定理说明了，相应于$\mathbf{C}^n$上的任一向量范数，都存在M_n上的一个相容矩阵范数. 该定理也给出了矩阵范数$\|\cdot\|$是由某个向量范数诱导的必要条件：$\| I \| = 1$；遗憾的是，这个必要条件并不是充分的.

下面，介绍几个重要的矩阵范数的例子，它们是由熟知的l_p范数诱导的，不过，不借助定义(5.6.1)也可以计算出来. 在以下每种情形，取$A = [a_{ij}] \in M_n$.

5.6.4 极大列和矩阵范数$\|\cdot\|_1$在M_n上定义为

$$\| A \|_1 \equiv \max_{1 \leqslant j \leqslant n} \sum_{i=1}^{n} | a_{ij} |.$$

范数$\|\cdot\|_1$是由l_1向量范数诱导的，因而它一定是矩阵范数. 可以把这一事实证明如下. 用A的各列把$A \in M_n$写成$A = [a_1 \cdots a_n]$. 于是$\| A \|_1 = \max_{1 \leqslant i \leqslant n} \| a_i \|_1$. 如果$x = [x_i]$，则

$$\|Ax\|_1 = \|x_1 a_1 + \cdots + x_n a_n\|_1 \leqslant \sum_{i=1}^{n} \|x_i a_i\|_1 = \sum_{i=1}^{n} |x_i| \|a_i\|_1$$

$$\leqslant \sum_{i=1}^{n} |x_i| \left(\max_{1\leqslant k\leqslant n} \|a_k\|_1 \right) = \sum_{i=1}^{n} |x_i| \, \|A\|_1 = \|x\|_1 \|A\|_1.$$

因而，$\max\limits_{\|x\|_1=1} \|Ax\|_1 \leqslant \|A\|_1$. 如果现在选 $x=e_k$（第 k 个单位基向量），则对任意 $k=1, 2, \cdots, n$ 有

$$\max_{\|x\|_1=1} \|Ax\|_1 \geqslant \|1a_k\|_1 = \|a_k\|_1,$$

因而

$$\max_{\|x\|_1=1} \|Ax\|_1 \geqslant \max_{1\leqslant k\leqslant n} \|a_k\|_1 \equiv \|A\|_1.$$

294

因为现在已经证明，由 l_1 向量范数诱导的矩阵范数既以$\|A\|_1$ 为上界又以$\|A\|_1$ 为下界，所以结论得证.

练习 试由定义直接证明$\|\cdot\|_1$ 是矩阵范数.

5.6.5 极大行和矩阵范数$\|\cdot\|_\infty$在 M_n 上定义为

$$\|A\|_\infty \equiv \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |a_{ij}|.$$

范数$\|\cdot\|_\infty$是由 l_∞向量范数诱导的，因而它一定是矩阵范数. 其证明与关于极大列和范数的证明类似. 算出

$$\|Ax\|_\infty = \max_{1\leqslant i\leqslant n} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \leqslant \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |a_{ij} x_j| \leqslant \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |a_{ij}| \, \|x\|_\infty = \|A\|_\infty \|x\|_\infty,$$

因而 $\max\limits_{\|x\|_\infty=1} \|Ax\|_\infty \leqslant \|A\|_\infty$. 如果 $A=0$，那就没有什么要证明的，所以可以假定 $A\neq 0$. 假定 A 的第 k 行非零，且定义向量 $z=[z_i]\in \mathbf{C}^n$ 为

$$\begin{cases} z_i = \dfrac{\bar{a}_{ki}}{|a_{ki}|}, \text{如果 } a_{ki} \neq 0; \\ z_i = 1, \text{如果 } a_{ki} = 0. \end{cases}$$

于是 $\|z\|_\infty=1$，且对所有 $j=1, 2, \cdots, n$，有 $a_{kj} z_j = |a_{kj}|$，并且

$$\max_{\|x\|_\infty=1} \|Ax\|_\infty \geqslant \|Az\|_\infty = \max_{1\leqslant i\leqslant n} \left| \sum_{j=1}^{n} a_{ij} z_j \right| \geqslant \left| \sum_{j=1}^{n} a_{kj} z_j \right| = \sum_{j=1}^{n} |a_{kj}|.$$

因此

$$\max_{\|x\|_\infty=1} \|Ax\|_\infty \geqslant \max_{1\leqslant k\leqslant n} \sum_{j=1}^{n} |a_{kj}| = \|A\|_\infty.$$

证毕.

练习 试由定义直接验证$\|\cdot\|_\infty$ 是 M_n 上的矩阵范数.

5.6.6 谱范数$\|\cdot\|_2$ 在 M_n 上定义为

$$\|A\|_2 \equiv \max\{\sqrt{\lambda} : \lambda \text{ 是 } A^* A \text{ 的特征值}\}.$$

295 注意，如果 $A^*Ax=\lambda x$ 且 $x\neq 0$，则 $x^*A^*Ax=\|Ax\|_2^2=\lambda\|x\|_2^2$，所以 $\lambda\geqslant 0$ 且 $\sqrt{\lambda}$ 是非负实数.

练习 如果 B 是正规矩阵，且 $B=V^*\Lambda V$，其中，U 是酉矩阵，且 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，证明

$$|x^*Bx|\leqslant \max\{|\lambda|: \lambda \text{ 是 } B \text{ 的特征值}\}\|x\|_2^2.$$

练习 证明 $\|Ax\|_2^2=x^*A^*Ax$ 对所有 $x\in\mathbf{C}^n$ 成立，然后利用上一个练习证明，$\|\cdot\|_2$ 是由 Euclid 向量范数 $\|\cdot\|_2$ 诱导的矩阵范数. 由此得出谱范数实际上是矩阵范数.

练习 证明 $\|UAV\|_2=\|A\|_2$ 对任意 $A\in M_n$ 和任意酉矩阵 U，$V\in M_n$ 成立. 因此，谱范数是酉不变矩阵范数.

下面证明，经一个固定的相似，一个矩阵范数可以变换成另一个矩阵范数.

5.6.7 定理 如果 $\|\cdot\|$ 是 M_n 上的矩阵范数，且 $S\in M_n$ 是非奇异矩阵，则对所有 $A\in M_n$，

$$\|A\|_S\equiv\|S^{-1}AS\|$$

是矩阵范数.

证明：可直接验证，$\|\cdot\|_S$ 适合公理(1)，(1a)，(2)和(3)，$\|\cdot\|_S$ 的次乘性可通过下面的计算得出：

$$\|AB\|_S=\|S^{-1}ABS\|=\|(S^{-1}AS)(S^{-1}BS)\|\leqslant\|S^{-1}AS\|\|S^{-1}BS\|=\|A\|_S\|B\|_S. \quad \square$$

为使矩阵范数适合特殊需要，定理(5.6.7)可能很有用. 这种形式的某些应用将在这里和下一节阐述.

矩阵范数的一个重要用途就是给出矩阵的谱的范围.

5.6.8 定义 矩阵 $A\in M_n$ 的谱半径 $\rho(A)$ 是

$$\rho(A)\equiv\max\{|\lambda|: \lambda \text{ 是 } A \text{ 的特征值}\}.$$

由此可知，如果 λ 是 A 的任一特征值，则 $|\lambda|\leqslant\rho(A)$；此外，至少有一个特征值 λ 可使 $|\lambda|=\rho(A)$. 设 $Ax=\lambda x$，$x\neq 0$，且 $|\lambda|=\rho(A)$，考虑其所有列都等于特征向量 x 的矩阵 $X\in M_n$，
296 并注意到 $AX=\lambda X$. 如果 $\|\cdot\|$ 是任意矩阵范数，则

$$|\lambda|\|X\|=\|\lambda X\|=\|AX\|\leqslant\|A\|\|X\|,$$

因此 $|\lambda|=\rho(A)\leqslant\|A\|$. 这便证明了下述定理.

5.6.9 定理 如果 $\|\cdot\|$ 是任意矩阵范数，且 $A\in M_n$，则 $\rho(A)\leqslant\|A\|$.

练习 试给出一个例子，使矩阵上的向量范数 $\|\cdot\|$ 和矩阵 $A\in M_n$ 适合 $\|A\|<\rho(A)$.

练习 设 $\|\cdot\|$ 是 M_n 上的矩阵范数，考虑映射 $F: \mathbf{C}^n\to M_n$，它定义为 $F(x)=[x\ x\ \cdots\ x]=M_n$ 中其所有列正好都是 x 的矩阵. 试证用 $\|x\|\equiv\|F(x)\|$ 定义的 $\mathbf{C}^n$ 上的函数 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上的范数，并且证明，对所有 $x\in\mathbf{C}^n$ 和所有 $A\in M_n$，有 $\|Ax\|\leqslant\|A\|\|x\|$. 这个不等式说明，向量范数 $\|\cdot\|$ 与矩阵范数 $\|\cdot\|$ 是相容的，而这个练习说明，M_n 上的任意矩阵范数在 $\mathbf{C}^n$ 上有一个相容向量范数.

虽然谱半径函数本身不是 M_n 上的矩阵或向量范数(见习题 19)，但是，对每个固定的 $A\in$

M_n，它是关于 A 的所有矩阵范数的值的最大下界.

5.6.10 引理 设 $A\in M_n$，且 $\varepsilon>0$ 是给定的. 则至少存在一个矩阵范数 $\|\cdot\|$ 使得 $\rho(A)\leqslant\|A\|\leqslant\rho(A)+\varepsilon$.

证明：根据 Schur 三角化定理(2.3.1)，存在酉矩阵 U 和上三角矩阵 Δ 使得 $A=U^*\Delta U$. 令 $D_t\equiv\mathrm{diag}(t, t^2, t^3, \cdots, t^n)$，由计算可知，

$$D_t\Delta D_t^{-1}=\begin{bmatrix}\lambda_1 & t^{-1}d_{12} & t^{-2}d_{13} & \cdots & t^{-n+1}d_{1n}\\ 0 & \lambda_2 & t^{-1}d_{23} & \cdots & t^{-n+2}d_{2n}\\ 0 & 0 & \lambda_3 & \cdots & t^{-n+3}d_{3n}\\ \vdots & \vdots & \vdots & \cdots & \vdots\\ 0 & 0 & 0 & \cdots & t^{-1}d_{n-1,n}\\ 0 & 0 & 0 & 0 & \lambda_n\end{bmatrix}$$

因此，对足够大的 $t>0$，可以确信，$D_t\Delta D_t^{-1}$ 的所有非对角元的绝对值的和小于 ε. 特别是，我们可以肯定，对足够大的 t 有 $\|D_t\Delta D_t^{-1}\|_1\leqslant\rho(A)+\varepsilon$. 这样，如果定义矩阵范数 $\|\cdot\|$ 为

$$\|B\|\equiv\|D_tU^*BUD_t^{-1}\|_1=\|(UD_t^{-1})^{-1}B(UD_t^{-1})\|_1,$$

297

其中 $B\in M_n$ 为任意矩阵，又如果选择足够大的 t，则可以构造出适合 $\|A\|\leqslant\rho(A)+\varepsilon$ 的矩阵范数. 因为 $\|A\|\geqslant\rho(A)$ 对任何矩阵范数成立，我们完成了证明. □

练习 说明为什么上述结果证明了 $\rho(A)=\inf\{\|A\|:\|\cdot\|$ 是矩阵范数$\}$.

我们的兴趣在于刻划当 $k\to\infty$ 时 $A^k\to0$ 的矩阵 A，下述结果是着手解决这个问题的最后一个工具.

5.6.11 引理 设 $A\in M_n$ 是给定的矩阵. 如果存在矩阵范数 $\|\cdot\|$ 使得 $\|A\|<1$，则 $\lim\limits_{k\to\infty}A^k=0$；即当 $k\to\infty$ 时 A^k 的所有元都趋于零.

证明：如果 $\|A\|<1$，则当 $k\to\infty$ 时有 $\|A^k\|\leqslant\|A\|^k\to0$. 这说明，关于范数 $\|\cdot\|$，$A^k\to0$，但是，因为 n^2 维空间 M_n 上的所有向量范数是等价的，所以，关于向量范数 $\|\cdot\|_\infty$，一定也有 $A^k\to0$. □

练习 试给一个例子，使矩阵 A 和矩阵范数 $\|\cdot\|_\alpha$，$\|\cdot\|_\beta$ 适合 $\|A\|_\alpha<1$ 和 $\|A\|_\beta>1$. 其结论是什么?

适合 $\lim\limits_{k\to\infty}A^k=0$ 的矩阵 $A\in M_n$ 称为收敛的，并且，在许多应用中，例如，在迭代过程的分析中，这样的矩阵是很有用的. 因此，重要的是能刻划收敛的矩阵.

5.6.12 定理 设 $A\in M_n$. 则 $\lim\limits_{k\to\infty}A^k=0$ 当且仅当 $\rho(A)<1$.

证明：如果 $A^k\to0$，又 $x\neq0$ 是适合 $Ax=\lambda x$ 的向量，则仅当 $|\lambda|<1$ 时才有 $A^kx=\lambda^kx\to0$. 因为 $|\lambda|<1$ 对 A 的每个特征值一定成立，得出 $\rho(A)<1$. 反过来，如果 $\rho(A)<1$，则根据引理(5.6.10)，存在某个矩阵范数 $\|\cdot\|$ 使得 $\|A\|<1$. 因此由引理(5.6.11)可知，当 $k\to\infty$ 时 $A^k\to0$. □

练习 考虑矩阵 $A=\begin{bmatrix}1/2 & 1\\ 0 & 1/2\end{bmatrix}\in M_2$. 对 $k=2, 3, \cdots$, 直接计算 A^k 和 $\rho(A^k)=[\rho(A)]^k$. 当$k\to\infty$时，下列各种情形如何变化? A^k 的各元; $\|A^k\|_1$; $\|A^k\|_\infty$; $\|A^k\|_2$.

298

练习 设 $A=\begin{bmatrix}1/2 & 1\\ -1/8 & 1/2\end{bmatrix}$, 且用递归式 $x^{(k+1)}=Ax^{(k)}$, $k=0, 1, \cdots$, 定义由向量$\{x^{(k)}\}\in\mathbf{C}^2$ 组成的序列. 证明，不论第一个向量 $x^{(0)}$ 如何选择，当 $k\to\infty$时 $x^{(k)}\to 0$.

有时，我们需要知道当 $k\to\infty$时，关于 A^k 的各元的大小的界. 一个有用的界是前一个定理的直接推论.

5.6.13 推论 设 $A\in M_n$ 是给定的矩阵，$\varepsilon>0$ 是给定的数. 则存在常数 $C=C(A, \varepsilon)$，使得

$$|(A^k)_{ij}|\leqslant C(\rho(A)+\varepsilon)^k$$

对所有 $k=1, 2, 3, \cdots$和所有 $i, j=1, 2, 3, \cdots, n$ 成立.

证明：因为矩阵 $\tilde{A}\equiv[\rho(A)+\varepsilon]^{-1}A$ 的谱半径严格小于 1，所以 $\tilde{A}$ 收敛，因而当 $k\to\infty$时 $\tilde{A}^k\to 0$. 特别是，序列$\{\tilde{A}^k\}$的各元有界，于是存在某个固定的 $C>0$，使得 $|(\tilde{A}^k)_{ij}|\leqslant C$ 对所有 $k=1, 2, 3, \cdots$和所有 $i, j=1, 2, \cdots, n$ 成立. 这就是所要确定的界. □

练习 设 $A=\begin{bmatrix}a & 1\\ 0 & a\end{bmatrix}$, 直接计算 A^k，说明在(5.6.13)中不能总取 $\varepsilon=0$.

当 $k\to\infty$时，虽然不能像 $\rho(A)^k$ 那样确切地说出 A^k 的各元的变化情况，然而，对任意矩阵范数$\|\cdot\|$，序列$\{\|A^k\|\}$的确具有下述渐近性质.

5.6.14 推论 设$\|\cdot\|$是 M_n 上的矩阵范数. 则对所有 $A\in M_n$，有

$$\rho(A)=\lim_{k\to\infty}\|A^k\|^{1/k}.$$

证明：因为 $\rho(A)^k=\rho(A^k)\leqslant\|A^k\|$，所以，对所有 $k=1, 2, \cdots$，有 $\rho(A)\leqslant\|A^k\|^{1/k}$. 如果 $\varepsilon>0$ 是给定的，则矩阵 $\tilde{A}\equiv[\rho(A)+\varepsilon]^{-1}A$ 的谱半径严格小于 1，因而 $\tilde{A}$ 收敛. 因此，当 $k\to\infty$时$\|\tilde{A}^k\|\to 0$，且存在某个 $N=N(\varepsilon, A)$，使得$\|\tilde{A}^k\|<1$ 对所有 $k\geqslant N$ 成立. 这正好说明，对所有 $k\geqslant N$，有$\|A^k\|\leqslant[\rho(A)+\varepsilon]^k$ 或$\|A^k\|^{1/k}\leqslant\rho(A)+\varepsilon$. 因为 $\rho(A)\leqslant\|A^k\|^{1/k}$对所有 k 成立，又因为 $\varepsilon>0$ 是任意的，由此得出$\lim\limits_{k\to\infty}\|A^k\|^{1/k}$存在且等于 $\rho(A)$. □

299

正像处理向量的无穷序列或无穷级数那样，也可以用向量范数处理有关矩阵的无穷序列或无穷级数的收敛性问题.

练习 设$\{A_k\}\subset M_n$ 是给定的矩阵无穷序列. 证明，如果在 M_n 上存在向量范数$\|\cdot\|$使得数值级数 $\sum\limits_{k=0}^{\infty}\|A_k\|$ 收敛(以致它的部分和有界)，则级数 $\sum\limits_{k=0}^{\infty}A_k$ 收敛于M_n 中的某个矩阵. 提示：证明其部分和构成一个 Cauchy 序列.

关于矩阵的一种特殊情形是矩阵的幂级数情形，这在研究向量的无穷级数时没有出现. 但是，因为矩阵范数的次乘性质，容易给出关于矩阵幂级数收敛性的简单的充分条件.

5.6.15 定理 设 $A\in M_n$，$\sum\limits_{k=0}^{\infty}a_kA^k$ 是无穷级数，如果存在 M_n 上的矩阵范数$\|\cdot\|$使得数值级数

$\sum_{k=0}^{\infty} |a_k| \|A\|^k$ 收敛，或者这个级数的部分和有界，则级数 $\sum_{k=0}^{\infty} a_k A^k$ 收敛.

练习 证明(5.6.15).

练习 用例子说明，有可能级数 $\sum_{k=0}^{\infty} a_k A^k$ 收敛，而级数 $\sum_{k=0}^{\infty} |a_k| \|A\|^k$ 发散. 这类似于数值级数的条件收敛(收敛而不绝对收敛).

练习 设函数 $f(z)$ 是用幂级数 $f(z) = \sum_{k=0}^{\infty} a_k z^k$ 定义的，它有收敛半径 $R>0$，且设 $\|\cdot\|$ 是 M_n 上的矩阵范数. 证明 $f(A) \equiv \sum_{k=0}^{\infty} a_k A^k$ 对所有适合 $\|A\|<R$ 的 $A \in M_n$ 有定义，更一般地，证明 $f(A)$ 对所有适合 $\rho(A)<R$ 的 $A \in M_n$ 有定义.

练习 如果 A 是可对角化的，且 $A=S^{-1}\Lambda S$，有时定义 $f(A) \equiv S^{-1} f(\Lambda) S$，其中 $f(\Lambda) \equiv \operatorname{diag}(f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n))$. 证明，如果 A 是可对角化的，则 $f(A)$ 的这个定义与上一个练习中的幂级数定义是一致的. 在这两个定义中，哪一个比较通用？

练习 证明用幂级数

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

给出的矩阵指数函数对每个 $A \in M_n$ 都有定义.

300

练习 应如何定义 $\cos(A)$？这对什么样的 A 有定义？

5.6.16 推论 设矩阵 $A \in M_n$，如果存在矩阵范数 $\|\cdot\|$ 使得 $\|I-A\|<1$，则 A 是可逆矩阵，且

$$A^{-1} = \sum_{k=0}^{\infty} (I-A)^k.$$

证明：如果 $\|I-A\|<1$，则因为级数 $\sum z^k$ 的收敛半径是 1，所以

$$\sum_{k=0}^{\infty} (I-A)^k$$

收敛于某个矩阵 C. 但是，因为 $N\to\infty$ 时，

$$A\sum_{k=0}^{N} (I-A)^k = [I-(I-A)]\sum_{k=0}^{N} (I-A)^k = I-(I-A)^{N+1} \to I,$$

所以，得出 $C=A^{-1}$. □

练习 证明上述结果等价于下述命题：如果 $\|\cdot\|$ 是矩阵范数，又如果 $\|A\|<1$，则 $I-A$ 是可逆矩阵，且

$$(I-A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

练习 设 $\|\cdot\|$ 是 M_n 上的矩阵范数，且假定给定的矩阵 $A \in M_n$ 有一个具有性质 $\|BA-I\|<1$ 的"近似逆" $B \in M_n$. 证明 A 和 B 都是可逆矩阵.

练习 如果矩阵范数 $\|\cdot\|$ 有 $\|I\|=1$ 的性质(如果它是诱导的范数，它理应是这样)，又如

果 $A\in M_n$ 适合 $\|A\|<1$．证明

$$\frac{1}{1+\|A\|}\leqslant\|(I-A)^{-1}\|\leqslant\frac{1}{1-\|A\|}.$$

提示：利用不等式 $\|(I-A)^{-1}\|\leqslant\sum_{k=0}^{\infty}\|A\|^k$ 可得到上界．对于下界，则利用一般的不等式 $\|B^{-1}\|\geqslant 1/\|B\|$ 和三角不等式．

练习 如果 $\|\cdot\|$ 是一般的矩阵范数，我们都知道 $\|I\|\geqslant 1$，在这种情形，证明，只要 $\|A\|<1$，就有

301

$$\frac{\|I\|}{\|I\|+\|A\|}\leqslant\|(I-A)^{-1}\|\leqslant\frac{\|I\|-(\|I\|-1)\|A\|}{1-\|A\|}.$$

练习 如果 A，$B\in M_n$，A 是可逆矩阵，又 $A+B$ 是奇异矩阵，证明 $\|B\|\geqslant 1/\|A^{-1}\|$ 对任何矩阵范数 $\|\cdot\|$ 成立．因此，可以用一个奇异矩阵去充分逼近一个非奇异矩阵，不过有一个内在的限度．**提示**：$A+B=A(I+A^{-1}B)$．如果 $\|A^{-1}B\|<1$，则 $I+A^{-1}B$ 应该是可逆的，因而有 $\|A^{-1}B\|\geqslant 1$.

关于可逆性的一个有用而又容易计算的准则不难由上一个推论得出．

5.6.17 推论 设 $A=[a_{ij}]\in M_n$，且假定对所有 $i=1,2,\cdots,n$ 有

$$|a_{ii}|>\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|,$$

则 A 是可逆矩阵．

证明：假设条件保证所有主对角元 a_{ii} 是非零的．令 $D=\mathrm{diag}(a_{11},\cdots,a_{nn})$，则 D 是可逆对角矩阵，$D^{-1}A$ 的主对角线上都是 1．$B=[b_{ij}]=I-D^{-1}A$ 的主对角线上都是零，且当 $i\neq j$ 时 $b_{ij}=-a_{ij}/a_{ii}$．考虑极大行和范数 $\|\cdot\|_\infty$．假设条件保证 $\|B\|_\infty<1$，于是根据(5.6.16)，$I-B=D^{-1}A$ 可逆，因而 A 可逆． □

满足推论(5.6.17)的矩阵称为*严格对角占优矩阵*．这个可逆性的充分条件称为 Levy-Desplanques定理，并且可以对它们作稍许改进．见(6.1)节，(6.2)节和(6.4)节．

现在更详细地讨论(5.6.1)中的诱导矩阵范数．这是一些最常见的矩阵范数，并且它们有一个重要的极小性质．因为常常需要采用检验 $\|A\|<1$ 的办法证明一个给定的矩阵 A 收敛，所以那些尽可能一致小的矩阵范数自然受到偏爱．正如要证明的那样，整个诱导矩阵范数类有这种合意的性质，并且这种性质刻化了诱导矩阵范数类．

有限维空间上的任意两种范数是等价的．所以，对每两种矩阵范数 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$，存在一个最小的有限正常数 $C_M(\alpha,\beta)$，使得 $\|A\|_\alpha\leqslant C_M(\alpha,\beta)\|A\|_\beta$ 对所有 $A\in M_n$ 成立，这个常数可以用

302

$$C_M(\alpha,\beta)=\max_{A\neq 0}\frac{\|A\|_\alpha}{\|A\|_\beta}$$

来计算．如果将 α 和 β 的作用颠倒过来，则一定存在一个定义类似的最小正常数 $C_M(\beta,\alpha)$，使

得$\|A\|_\beta \leqslant C_M(\beta, \alpha)\|A\|_\alpha$对所有$A\in M_n$成立．一般地，两个常数$C_M(\alpha, \beta)$和$C_M(\beta, \alpha)$之间没有明显的关系，不过，如果我们考察本节末习题23中的表，它的左上角的3×3数表是对称的：即$C_M(\alpha, \beta)=C_M(\beta, \alpha)$对三个矩阵范数$\|\cdot\|_1$，$\|\cdot\|_2$和$\|\cdot\|_\infty$中的任意一对都成立．这三个矩阵范数都是诱导范数，而上述对称性是所有诱导范数的一个性质．

5.6.18 定理 设$\|\cdot\|_\alpha$和$\|\cdot\|_\beta$是$\mathbf{C}^n$上两个给定的向量范数，且设$\|\cdot\|_\alpha$和$\|\cdot\|_\beta$表示M_n上的相应诱导矩阵范数，即

$$\|A\|_\alpha \equiv \max_{x\neq 0}\frac{\|Ax\|_\alpha}{\|x\|_\alpha} \text{ 和 } \|A\|_\beta \equiv \max_{x\neq 0}\frac{\|Ax\|_\beta}{\|x\|_\beta}.$$

定义

$$R_{\alpha\beta} \equiv \max_{x\neq 0}\frac{\|x\|_\alpha}{\|x\|_\beta} \text{ 和 } R_{\beta\alpha} \equiv \max_{x\neq 0}\frac{\|x\|_\beta}{\|x\|_\alpha} \tag{5.6.19}$$

则

$$\max_{A\neq 0}\frac{\|A\|_\alpha}{\|A\|_\beta} = R_{\alpha\beta}R_{\beta\alpha}. \tag{5.6.20}$$

特别是

$$\max_{A\neq 0}\frac{\|A\|_\alpha}{\|A\|_\beta} = \max_{A\neq 0}\frac{\|A\|_\beta}{\|A\|_\alpha} = R_{\alpha\beta}R_{\beta\alpha}. \tag{5.6.21}$$

证明：设$A\in M_n$和$x\in\mathbf{C}^n$是给定的，且假定$x\neq 0$和$Ax\neq 0$．则

$$\frac{\|Ax\|_\alpha}{\|x\|_\alpha} = \frac{\|Ax\|_\alpha}{\|Ax\|_\beta}\frac{\|Ax\|_\beta}{\|x\|_\beta}\frac{\|x\|_\beta}{\|x\|_\alpha} \leqslant R_{\alpha\beta}\frac{\|Ax\|_\beta}{\|x\|_\beta}R_{\beta\alpha},$$

并且，即使$Ax=0$，这个不等式也成立．于是

$$\|A\|_\alpha = \max_{x\neq 0}\frac{\|Ax\|_\alpha}{\|x\|_\alpha} \leqslant R_{\alpha\beta}\max_{x\neq 0}\frac{\|Ax\|_\beta}{\|x\|_\beta}R_{\beta\alpha} = R_{\alpha\beta}R_{\beta\alpha}\|A\|_\beta,$$

因而对所有非零$A\in M_n$有

303

$$\frac{\|A\|_\alpha}{\|A\|_\beta} \leqslant R_{\alpha\beta}R_{\beta\alpha}. \tag{5.6.22}$$

(5.6.19)中的每个极值都被某个非零向量所达到．所以存在向量y，$z\in\mathbf{C}^n$，使得$\|y\|_2=\|z\|_2=1$，$\|y\|_\alpha=R_{\alpha\beta}\|y\|_\beta$和$\|z\|_\beta=R_{\beta\alpha}\|z\|_\alpha$．根据推论(5.5.15)，存在向量$z_0\in\mathbf{C}^n$使得

(a) $|z_0^*x|\leqslant\|x\|_\beta$对所有$x\in\mathbf{C}^n$成立；

(b) $z_0^*z=\|z\|_\beta$．

考虑矩阵$A_0\equiv yz_0^*$．利用(b)有

$$\frac{\|A_0z\|_\alpha}{\|z\|_\alpha} = \frac{\|yz_0^*z\|_\alpha}{\|z\|_\alpha} = \frac{\|y\|_\alpha|z_0^*z|}{\|z\|_\alpha} = \frac{\|y\|_\alpha\|z\|_\beta}{\|z\|_\alpha},$$

所以有下界

$$\|A_0\|_\alpha \geqslant \frac{\|y\|_\alpha\|z\|_\beta}{\|z\|_\alpha} = R_{\alpha\beta}R_{\beta\alpha}\|y\|_\beta.$$

另一方面，可以利用(a)得到

$$\frac{\|A_0 x\|_\beta}{\|x\|_\beta}=\frac{\|yz_0^* x\|_\beta}{\|x\|_\beta}=\frac{\|y\|_\beta\,|z_0^* x|}{\|x\|_\beta}\leqslant\frac{\|y\|_\beta\,\|x\|_\beta}{\|x\|_\beta}=\|y\|_\beta,$$

因而有上界

$$\|A_0\|_\beta\leqslant\|y\|_\beta.$$

合并这两个界便有

$$\frac{\|A_0\|_\alpha}{\|A_0\|_\beta}\geqslant\frac{R_{\alpha\beta}R_{\beta\alpha}\|y\|_\beta}{\|y\|_\beta}=R_{\alpha\beta}R_{\beta\alpha},$$

它说明(5.6.22)中等式可以成立，因而证明了(5.6.20)．因为恒等式(5.6.20)的右边关于 α 和 β 是对称的．所以结论(5.6.21)成立． □

$\mathbf{C}^n$ 上的两个不同的向量范数能够诱导出 M_n 上的相同矩阵范数吗？根据(5.6.18)的下述推论可知，这种情形能够出现，当且仅当一个向量范数是另一个向量范数的常纯量倍．

5.6.23 推论 设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 $\mathbf{C}^n$ 上的向量范数．又设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 表示 M_n 上的相应的诱导矩阵范数．则 $\|A\|_\alpha=\|A\|_\beta$ 对所有 $A\in M_n$ 成立，当且仅当存在一个正常数 c，使得 $\|x\|_\alpha=$ 304 $c\|x\|_\beta$ 对所有 $x\in\mathbf{C}^n$ 成立．

证明： 我们知道．

$$R_{\beta\alpha}=\max_{x\neq0}\frac{\|x\|_\beta}{\|x\|_\alpha}=\left[\min_{x\neq0}\frac{\|x\|_\alpha}{\|x\|_\beta}\right]^{-1}\geqslant\left[\max_{x\neq0}\frac{\|x\|_\alpha}{\|x\|_\beta}\right]^{-1}=\frac{1}{R_{\alpha\beta}}.$$

因此，有一般不等式

$$R_{\alpha\beta}R_{\beta\alpha}\geqslant1,\tag{5.6.24}$$

其中等式成立．当且仅当

$$\min_{x\neq0}\frac{\|x\|_\alpha}{\|x\|_\beta}=\max_{x\neq0}\frac{\|x\|_\alpha}{\|x\|_\beta},$$

而这可以成立．当且仅当对所有 $x\neq0$．函数 $\|x\|_\alpha/\|x\|_\beta$ 是常数．因此，如果 $\|x\|_\alpha=c\|x\|_\beta$，就一定有 $R_{\alpha\beta}R_{\beta\alpha}=1$，于是，由(5.6.21)可知．对所有 $A\in M_n$，$\|A\|_\alpha\leqslant\|A\|_\beta$，并且 $\|A\|_\beta\leqslant\|A\|_\alpha$；因而．对所有 $A\in M_n$，$\|A\|_\alpha=\|A\|_\beta$．反过来，如果两个诱导矩阵范数相同，则由(5.6.20)可知 $R_{\alpha\beta}R_{\beta\alpha}=1$，因而在(5.6.24)中等式成立．由前面的证明得知，比值 $\|x\|_\alpha/\|x\|_\beta$ 是常数． □

5.6.25 推论 设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 $\mathbf{C}^n$ 上的向量范数，又设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 表示 M_n 上的相应诱导矩阵范数．则 $\|A\|_\alpha\leqslant\|A\|_\beta$ 对所有 $A\in M_n$ 成立．当且仅当 $\|A\|_\alpha=\|A\|_\beta$ 对所有 $A\in M_n$ 成立．

证明： 如果 $\|A\|_\alpha\leqslant\|A\|_\beta$ 对所有 $A\in M_n$ 成立．则有 $R_{\alpha\beta}R_{\beta\alpha}\leqslant1$，[因为(5.6.24)]这蕴涵 $R_{\alpha\beta}R_{\beta\alpha}=1$．因此，根据(5.6.21)，对所有 $A\in M_n$，$\|A\|_\alpha\leqslant\|A\|_\beta$ 且 $\|A\|_\beta\leqslant\|A\|_\alpha$． □

最后一个推论说明，没有一个诱导矩阵范数能够一致地小于另一个诱导范数．如果允许它同其他(不一定是诱导的)矩阵范数相比较．会出现什么情形呢？

5.6.26 定理 设 $\|\cdot\|$ 是 M_n 上给定的矩阵范数，且设 $\|\cdot\|_\alpha$ 是 M_n 上给定的诱导矩阵范数．则

(a) 存在 M_n 上的诱导矩阵范数 $N(\cdot)$ 使得对每个 $A\in M_n$ 有 $N(A)\leqslant\|A\|$；

(b) $\|A\| \leqslant \|A\|_\alpha$ 对每个 $A\in M_n$ 成立，当且仅当 $\|A\| = \|A\|_\alpha$ 对每个 $A\in M_n$ 成立.

证明：定义 $\mathbf{C}^n$ 上的向量范数 $\|\cdot\|$ 为

$$\|x\| = \|X\|, \quad X = [x\ x\ \cdots\ x] \in M_n, \tag{5.6.27}$$

[305]

并且考虑由 $\|\cdot\|$ 诱导的 M_n 上的矩阵范数 $N(\cdot)$. 对任意 $A\in M_n$，有

$$N(A) = \max_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \max_{x\neq 0} \frac{\|[Ax\ Ax \cdots Ax]\|}{\|[x x \cdots x]\|} = \max_{x\neq 0} \frac{\|AX\|}{\|X\|}$$

$$\leqslant \max_{x\neq 0} \frac{\|A\|\|X\|}{\|X\|} = \|A\| \quad (\text{因为}\|\cdot\|\text{是矩阵范数}), \tag{5.6.28}$$

这就证明了(a). 为了证明(b)，假定 $\|A\| \leqslant \|A\|_\alpha$ 对所有 $A\in M_n$ 成立. 则由(a)可知，对所有 $A\in M_n$ 有

$$N(A) \leqslant \|A\| \leqslant \|A\|_\alpha.$$

但是 $N(\cdot)$ 和 $\|\cdot\|_\alpha$ 都是诱导范数. 所以由(5.6.25)可知，$N(A) = \|A\|_\alpha$，因而对所有 $A\in M_n$ 有 $\|A\| = \|A\|_\alpha$. □

上述结果是促成给出下述定义的原因.

5.6.29 定义 设 $\|\cdot\|$ 是 M_n 上的矩阵范数，如果对所有 $A\in M_n$，适合 $N(A) \leqslant \|A\|$ 的 M_n 上的矩阵范数只有 $N(\cdot) = \|\cdot\|$，就称 $\|\cdot\|$ 是 M_n 上的极小矩阵范数.

定理(5.6.26)的论断(b)说明，M_n 的每个诱导范数是极小的. 论断(a)直接推出每个极小范数是诱导范数. 因此，如果想采用一个矩阵范数，而它又不能(用所有矩阵上的小值)一致地加以改进，那么就可以采用诱导范数，并且，具有这种极小性的任一范数一定是诱导范数.

向量范数(5.6.27)是可由一个给定的矩阵范数构造出来的整个向量范数族的特殊情形. 设 $\|\cdot\|$ 是 M_n 上给定的矩阵范数，$y\in \mathbf{C}^n$ 是给定的非零向量，且用

$$\|x\|_y = \|xy^*\|, \quad y\in \mathbf{C}^n, \quad y\neq 0 \tag{5.6.30}$$

定义函数 $\|x\|_y$：$\mathbf{C}^n \to \mathbf{R}$，则 $\|\cdot\|_y$ 是 $\mathbf{C}^n$ 上的向量范数，且对所有 $A\in M_n$ 具有性质

$$\|Ax\|_y = \|A(xy^*)\| \leqslant \|A\|\,\|xy^*\| = \|A\|\,\|x\|_y.$$

如果 $y = [1\ 1\ \cdots\ 1]^T$，则(5.6.30)简化成(5.6.27). 如果用 $N_y(\cdot)$ 表示由 $\|\cdot\|_y$ 诱导的 M_n 上的矩阵范数，这个不等式说明，对所有 $A\in M_n$，有

[306]

$$N_y(A) = \max_{x\neq 0} \frac{\|Ax\|_y}{\|x\|_y} \leqslant \max_{x\neq 0} \frac{\|A\|\,\|x\|_y}{\|x\|_y} = \|A\|. \tag{5.6.31}$$

这显然是(5.6.26a)的推广.

如果给定的矩阵范数 $\|\cdot\|$ 是极小范数，则(5.6.31)说明，对所有 $A\in M$，$\|A\| = N_y(A)$. 因为在上述论证中所采用的向量 y 可以是任意非零向量，于是，对所有非零 y，$z\in \mathbf{C}^n$，应有 $N_y(\cdot) = \|\cdot\| = N_z(\cdot)$.

5.6.32 定理 设 $\|\cdot\|$ 是 M_n 上的矩阵范数，$N_y(\cdot)$ 是用(5.6.31)和(5.6.30)定义的诱导范数. 则下列命题等价：

(a) $\|\cdot\|$是诱导矩阵范数.

(b) $\|\cdot\|$是极小矩阵范数.

(c) $\|\cdot\|=N_y(\cdot)$对所有非零 $y\in\mathbf{C}^n$ 成立.

证明： (a)蕴涵(b)的论断正是(5.6.26b). 我们刚才已经看到，如果$\|\cdot\|$是极小范数，则$\|\cdot\|=N_y(\cdot)$，所以(b)蕴涵(c). 如果(c)成立，则$\|\cdot\|$是诱导范数，因为根据定义，$N_y(\cdot)$是诱导范数. □

从这些论断还可得出一些结果. 如果 $N_y(\cdot)=\|\cdot\|$对所有非零 $y\in\mathbf{C}^n$ 成立，则 $N_y(\cdot)$对所有非零 $y,z\in\mathbf{C}^n$ 成立. 但是推论(5.6.23)说明，除了差一个纯量因子以外，诱导一个给定的矩阵范数的向量是唯一确定的，所以，对某个正常数 c_{yz}，有$\|\cdot\|_y=c_{yz}\|\cdot\|_z$.

练习　如果 M_n 上的矩阵范数$\|\cdot\|$是由 $\mathbf{C}^n$ 上的向量范数$\|\cdot\|$诱导的，证明，对所有 $y,z\in\mathbf{C}^n$，有$\|yz^*\|=\|y\|\|z\|^D$，$\|\cdot\|_z=\|\cdot\|\|z\|^D$ 以及 $c_{yz}=\|y\|^D/\|z\|^D$. 向量范数$\|\cdot\|^D$如(5.4.12)中定义的那样，是向量范数$\|\cdot\|$的对偶.

5.6.33 定理　设$\|\cdot\|$是 M_n 上给定的矩阵范数，且设$\|\cdot\|_y$是用(5.6.30)定义的 $\mathbf{C}^n$ 上的向量范数. 则下列两个论断等价：

(a) 对每对非零向量 $y,z\in\mathbf{C}^n$，有正常数 c_{yz} 使$\|x\|_y=c_{yz}\|x\|_z$对所有 $x\in\mathbf{C}^n$ 成立.

307　(b) $\|xy^*\|=\dfrac{\|xz^*\|\|zy^*\|}{\|zz^*\|}$对所有 $x,y,z\in\mathbf{C}^n$ 且 $z\neq 0$ 成立.

如果$\|\cdot\|$是诱导矩阵范数，则它满足恒等式(b)，且由它通过(5.6.30)构造的向量范数满足(a).

证明： 如果(a)成立，则

$\|xz^*\|\|zy^*\|=\|x\|_z\|z\|_y=(1/c_{yz})\|x\|_y c_{yz}\|z\|_z=\|x\|_y\|z\|_z=\|xy^*\|\|zz^*\|$.

反过来，如果(b)成立，则(a)成立，且 $c_{yz}=\|zy^*\|/\|zz^*\|$. 已经证明，如果 $N_y(\cdot)=\|\cdot\|$，则(a)[因而(b)也]一定成立，而如果$\|\cdot\|$是诱导范数，则根据(5.6.32)，情况就是这样. □

练习　我们知道，一个诱导范数的任一正纯量倍数满足恒等式(5.6.33b). 证明矩阵范数$\|\cdot\|_1$和$\|\cdot\|_2$都满足这个恒等式，不过这两个范数都不是一个诱导范数的纯量倍数.

在(5.6.2)中已经看到，如果$\|\cdot\|$是诱导范数，则$\|I\|=1$. 遗憾的是，这个性质对矩阵范数为诱导范数的事实不是充分的. 容易证明，函数

$$\|A\|\equiv\max\{\|A\|_1,\|A\|_\infty\}\tag{5.6.34}$$

定义了 M_n 上的一个矩阵范数，且$\|I\|=1$. 但是，因为$\|A\|_1\leqslant\|A\|$对所有 $A\in M_n$ 成立，且当 $A=\begin{bmatrix}1&0\\1&3\end{bmatrix}$时，$\|A\|_1<\|A\|$，所以$\|\cdot\|$不是极小范数，因而不可能是诱导范数.

练习　验证(5.6.34)定义了一个矩阵范数. 更一般地，证明，如果$\|\cdot\|_{(1)},\cdots,\|\cdot\|_{(k)}$是 M_n 上给定的矩阵范数，则

$$\|A\|=\max\{\|A\|_{(1)},\cdots,\|A\|_{(k)}\}$$

定义了 M_n 上的一个矩阵范数.

诱导范数在所有矩阵范数中是极小范数，但是，假定现在只考虑由酉不变矩阵范数组成的一类重要范数. 这是一些对所有 $A\in M_n$ 和所有酉矩阵 U，$V\in M_n$，适合 $\|A\|=\|UAV\|$ 的矩阵范数 $\|\cdot\|$. 可以证明，在这类范数中只有一个极小范数，那就是谱范数.

5.6.35 推论 如果 $\|\cdot\|$ 是酉不变矩阵范数，则 $\|A\|_2\leqslant\|A\|$ 对所有 $A\in M_n$ 成立. 谱范数 $\|\cdot\|_2$ 是 M_n 上仅有的既为诱导的又为酉不变的矩阵范数.

证明： 假定 $\|\cdot\|$ 是给定的酉不变矩阵范数，由定理(5.6.26)的(a)可知，$N(A)\leqslant\|A\|$ 对所有 $A\in M_n$ 成立，其中 $N(A)$ 是由用(5.6.27)定义的向量范数 $\|\cdot\|$ 诱导的范数，如果 $U\in M_n$ 是酉矩阵，则有 $\|Ux\|=\|UX\|=\|X\|=\|x\|$，因而向量范数 $\|\cdot\|$ 是酉不变的. 如果 $x\in\mathbf{C}^n$ 是给定的非零向量，则存在酉矩阵 U 使得 $Ux=\|x\|_2e_1$. 于是对所有 $x\in\mathbf{C}^n$ 有 $\|x\|=\|\|x\|_2U^*e_1\|=\|x\|_2\|U^*e_1\|=\|x\|_2\|e_1\|$. 因而，向量范数 $\|\cdot\|$ 是 Euclid 范数的纯量倍，而推论(5.6.23)说明，(用 $\|\cdot\|$ 诱导的矩阵范数) $N(\cdot)$ 等于(用 $\|\cdot\|_2$ 诱导的矩阵范数) $\|\cdot\|_2$. 因此对所有 $A\in M_n$ 有 $\|A\|_2=N(A)\leqslant\|A\|$. 如果假定 $\|\cdot\|$ 是诱导范数，则它是极小的，因而 $\|A\|_2=\|A\|$ 对所有 $A\in M_n$ 成立. □

如果 $\|\cdot\|$ 是 M_n 上的矩阵范数，利用

$$\|A\|^*\equiv\|A^*\|$$

定义的函数 $\|\cdot\|^*$ 也是 M_n 上的矩阵范数，直接计算说明，对所有 $A\in M_n$，$\|A\|_2^*=\|A^*\|_2=\|A\|_2$，且 $\|A\|_1^*=\|A^*\|_1=\|A\|_\infty$，但是，并非每个矩阵范数都有这个性质，这是因为 $\|A\|_1^*=\|A\|_\infty\neq\|A\|_1$. 适合 $\|\cdot\|^*=\|\cdot\|$ 的矩阵范数称为自伴的. Frobenius 矩阵范数和 l_1 矩阵范数是自伴的，又因为

$$\|A^*\|_2^2=\rho(AA^*)=\rho(A^*A)=\|A\|_2^2,$$

所以谱范数也是自伴的. 实际上，M_n 上的所有酉不变范数都是自伴的[见(7.4)节，习题 2]. 谱范数可看作是仅有的自伴诱导矩阵范数.

5.6.36 定理 设 $\|\cdot\|$ 是 M_n 上给定的矩阵范数. 则

(a) $\|\cdot\|^*$ 是诱导范数，当且仅当 $\|\cdot\|$ 是诱导范数.

(b) 如果矩阵范数 $\|\cdot\|$ 是由向量范数 $\|\cdot\|$ 诱导的，则 $\|\cdot\|^*$ 是由对偶范数 $\|\cdot\|^D$ 诱导的.

(c) 谱范数是 M_n 上仅有的既为诱导的又为自伴的矩阵范数.

证明： 如果 $N(\cdot)$ 是矩阵范数，又如果 $N(A)\leqslant\|A\|^*=\|A^*\|$ 对所有 $A\in M_n$ 成立，则 $N(A^*)^*=N(A^*)\leqslant\|A\|$ 对所有 $A\in M_n$ 成立. 如果 $\|\cdot\|$ 是极小矩阵范数，$N(\cdot)^*=\|\cdot\|$，因而 $N(\cdot)=\|\cdot\|^*$，所以 $\|\cdot\|^*$ 是极小矩阵范数. 由(5.6.32)可知论断(a)成立. 现在假定 $\|\cdot\|$ 是由向量范数 $\|\cdot\|$ 诱导的. 利用对偶性定理(5.5.14)，有

$$\|A\|^*=\|A^*\|=\max_{\|x\|=1}\|A^*x\|=\max_{\|x\|=1}(\|A^*x\|^D)^D$$

$$=\max_{\|x\|=1}\max_{\|z\|^D=1}|(A^*x)^*z|=\max_{\|z\|^D=1}\max_{\|x\|=1}|x^*Az|=\max_{\|z\|^D=1}\|Az\|^D,$$

因而$\|\cdot\|$是由$|\cdot|^D$诱导的. 关于最后一个论断，我们注意到，如果矩阵范数$\|\cdot\|$是由向量范数$|\cdot|$诱导的，且$\|\cdot\|=\|\cdot\|^*$，(b)说明$\|\cdot\|$也是由$|\cdot|^D$诱导的. 但是推论(5.6.23)说明，除了差一个正纯量因子外，诱导一个给定的矩阵范数的向量范数是唯一确定的，因而，存在某个$c>0$使得$|\cdot|^D=c|\cdot|$. 于是，由(5.4.16)，一定有$|\cdot| = \|\cdot\|_2\sqrt{c}$. 因为给定的向量范数是Eucldi范数的一个倍数，所以它们都诱导同一个矩阵范数，由此得出$\|\cdot\|=\|\cdot\|_2$. □

练习　证明，只要$\|\cdot\|$是矩阵范数，$\|\cdot\|^*$就是矩阵范数.

练习　给出一个例子，说明自伴矩阵范数不一定是酉不变范数.

在(5.5)中引进的绝对向量范数和单调向量范数是最通用的向量范数. 用单调向量范数诱导的矩阵范数有一个简单而又有用的特征.

5.6.37 定理　设$|\cdot|$是$\mathbf{C}^n$上的向量范数，$\|\cdot\|$是由它诱导的M_n上的矩阵范数，则下列命题等价：

(a) $|\cdot|$是绝对范数；即$|\,|x|\,| = |x|$对所有$x\in\mathbf{C}^n$成立.

(b) $|\cdot|$是单调范数；即只要$|x|\leqslant|y|$，就有$|x|\leqslant|y|$.

(c) 如果$D=\operatorname{diag}(d_1, d_2, \cdots, d_n)\in M_n$，则

$$\|D\| = \max_{1\leqslant i\leqslant n}|d_i|.$$

证明：(a)与(b)等价是(5.5.10)的内容. 如果$|\cdot|$是单调范数，又如果令

$$d \equiv \max_{1\leqslant i\leqslant n}|d_i|, d=|d_k|,$$

则$|Dx|\leqslant|dx|$，因而$|Dx|\leqslant d|x|$，且$x=e_k$时等式成立. 于是

$$\|D\| = \max_{x\neq 0}\frac{|Dx|}{|x|} = d$$

因而(b)蕴涵(c). 如果假定(c)成立，设$x, y\in\mathbf{C}^n$是给定的，且$|x|\leqslant|y|$，注意到存在一组复数d_k使得$|x_k|=d_ky_k$，且$|d_k|\leqslant1$，$k=1, \cdots, n$. 因此，如果$D=\operatorname{diag}(d_1, \cdots, d_n)$，则有$Dy=|x|$和$\|D\|\leqslant1$. 因为

[310]

$$|\,|x|\,| = |Dy| \leqslant \|D\|\,|y| \leqslant |y|,$$

所以范数$|\cdot|$一定是单调的. □

习题

1. 试给一个关于矩阵的向量范数的例子，它适合$\|I\|<1$.

2. 满足$A^2=A$的矩阵A称的为幂等的. 试给一个不同于I和0的2×2幂等矩阵的例子. 证明0和1是幂等矩阵仅有的特征值. 证明幂等矩阵A一定可对角化，又如果$A\neq0$，则$\|A\|\geqslant1$对任何矩阵范数成立.

3. 如果$\|\cdot\|$是M_n上的矩阵范数，证明，对所有$c\geqslant1$，$c\|\cdot\|$是矩阵范数. 但是证明，对任意$c<1$，$c\|\cdot\|_1$和$c\|\cdot\|_1$都不是矩阵范数.

4. 在定义(5.6.1)中，同一种向量范数有两种不同的计算方式. 更一般地，可以定义$\|\cdot\|_{\alpha,\beta}$为

$$\|A\|_{\alpha,\beta} \equiv \max_{\|x\|_\alpha=1}\|Ax\|_\beta$$

其中 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是两个(可能不同的)向量范数. 这样的函数 $\|\cdot\|_{\alpha,\beta}$ 是矩阵范数吗? 试研究一下 $\|\cdot\|_{\alpha,\beta}$, 确定它可能具有那些有趣的性质: 注意, 这个概念可用来定义关于 $m\times n$ 矩阵的范数, 这是因为 $\|\cdot\|_\alpha$ 可以取 $\mathbf{C}^m$ 上的向量范数, 而 $\|\cdot\|_\beta$ 可以取 $\mathbf{C}^n$ 上的向量范数. 在这个意义下, $\|\cdot\|_{\alpha,\beta}$ 有哪些性质与诱导矩阵范数是一样的?

5. 证明 Euclid 范数 $\|\cdot\|_2$ 和谱范数 $\|\cdot\|_2$ 都是 M_n 上的酉不变范数; 也就是说, 只要 U 和 V 都是酉矩阵, A 与 UAV 就有相同的范数. 试就你能考虑到的各个方面, 对矩阵范数 $\|\cdot\|_2$ 与 $\|\cdot\|_2$ 作一比较. 注意 $\|A\|_2=[\operatorname{tr} A^*A]^{1/2}$.

6. 证明关于 $\|\cdot\|$ 的公理(1)~(3)对(5.6.7)中的 $\|\cdot\|_S$ 同样成立. 这证明了, 如果在(5.6.7)的假设和结论中用"关于矩阵的向量范数"代替"矩阵范数", (5.6.7)仍然成立.

7. 如果 $\|\cdot\|$ 是 M_n 上的诱导矩阵范数, 又如果 $S\in M_n$ 是非奇异矩阵, 证明[如(5.6.7)中定义的] $\|\cdot\|_S$ 也是诱导矩阵范数. 如果 $\|\cdot\|$ 是由向量范数 $\|\cdot\|$ 诱导的, 证明矩阵范数 $\|\cdot\|_S$ 是[如(5.3.2)中定义的]向量范数 $\|\cdot\|_S$ 诱导的.

8. 证明 M_n 的非奇异矩阵在 M_n 中是稠密的: 即证明 M_n 中的每个矩阵是诸非奇异矩阵的极限, 奇异矩阵也有 M_n 中稠密吗?

9. 证明, 对所有 $m\geqslant 1$, 由 $\mathbf{C}^m$ 上的向量范数组成的集合是凸集. 但是, 对任意 $n\geqslant 2$, 由 M_n 上的矩阵范数组成的集合不是凸集. 证明, $N(\cdot)=\frac{1}{2}[N_1(\cdot)+N_2(\cdot)]$ 是矩阵范数, 当且仅当对所有 A, $B\in M_n$ 有

$$
\begin{aligned}
[N_1(A)-N_2(A)][N_1(B)-N_2(B)] \leqslant\ & 2[N_1(A)N_1(B)-N_1(AB)] \\
& +2[N_2(A)N_2(B)-N_2(AB)].
\end{aligned}
$$

提示: 考虑 $N_1(\cdot)=\|\cdot\|_1$, $N_2(\cdot)=\|\cdot\|_2$, $A=\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$ 及 $B=A^T$. 关于矩阵范数集的一个重要子集是凸集的事实见例(7.4.54).

10. 证明, M_n 上的 l_1 范数 $\|A\|_1=\sum_{i,j=1}^{n}|a_{ij}|$ 是矩阵范数, 但不是诱导范数.

11. 证明下述各个恒等式都是计算谱范数(5.6.6)的等价形式:

$$
\begin{aligned}
\|A\|_2 &= \max_{\|x\|_2=1}\|Ax\|_2 = \max_{\|x\|_2\leqslant 1}\|Ax\|_2 = \max_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2} \\
&= \max_{\|x\|_2=\|y\|_2=1}|y^*Ax| = \max_{\substack{\|x\|_2\leqslant 1\\ \|y\|_2\leqslant 1}}|y^*Ax|.
\end{aligned}
$$

利用这些恒等式证明 $\|A\|_2=\|A^*\|_2$ 对所有 $A\in M_n$ 成立. 然后利用 $\|\cdot\|_2$ 是矩阵范数和 A^*A 是 Hermite 矩阵的事实证明 $\|AA^*\|_2=\|A^*A\|_2=\|A\|_2^2$.

12. 如果 $\rho(A)<1$, $A\in M_n$, 证明级数 $I+A+A^2+\cdots$ 收敛于和 $(I-A)^{-1}$.

13. 如果 $A\in M_n$ 不是可逆矩阵, 证明对每个矩阵范数 $\|\cdot\|$ 有 $\|I-A\|\geqslant 1$.

14. 设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 M_n 上给定的矩阵范数. 证明 $\|A\|\equiv\max\{\|A\|_\alpha, \|A\|_\beta\}$ 是 M_n 上的矩阵范数. 什么时候它是诱导范数?

311
312

15. 试给一个矩阵 A 的例子，使得 $\rho(A)<|||A|||$ 对每个矩阵范数 $|||\cdot|||$ 成立.

16. 设 $A=[a_{ij}]\in M_n$. 证明，用 $|||A|||\equiv n\max\limits_{1\leq i,j\leq n}|a_{ij}|$ 定义的 M_n 上的函数 $|||\cdot|||$ 是矩阵范数，而当 $n\geq 2$ 时，它不是诱导范数.

17. 利用习题 12 的想法计算矩阵

$$\begin{bmatrix}1 & 2 & 1\\ 0 & 1 & 3\\ 0 & 0 & 1\end{bmatrix}$$

的逆. **提示**：在级数中只有三项非零.

18. 说明如何推广习题 17 中的方法求一般非奇异上三角矩阵 $A\in M_n$ 的逆. **提示**：选一个对角矩阵 D，使得 DA 的主对角线上都是 1.

19. 证明，谱半径 $\rho(\cdot)$ 是 M_n 上的连续齐次函数，但是它既不是 M_n 上矩阵范数，也不是 M_n 上的向量范数，这是因为，

(a) 可能对某个 $A\neq 0$，有 $\rho(A)=0$；

(b) 可能有 $\rho(A+B)>\rho(A)+\rho(B)$；

(c) 即使 $\rho(A)$ 和 $\rho(B)$ 都是非零的，也可能有 $\rho(AB)>\rho(A)\rho(B)$.

提示：考虑 $\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$，$\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}$，$\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$ 和 $\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$.

20. 证明 $\|AB\|_2\leq|||A|||_2\|B\|_2$ 和 $\|AB\|_2\leq\|A\|_2|||B|||_2$ 对所有 A，$B\in M_n$ 成立.

21. 证明 $|||A|||_2^2\leq|||A|||_1|||A|||_\infty$ 对所有 $A\in M_n$ 成立. **提示**：$\rho(A^*A)\leq|||A^*A|||_1$，且 $|||A^*|||_1=|||A|||_\infty$.

22. 设 $\|\cdot\|_\alpha$ 是 $\mathbf{C}^n$ 上给定的向量范数，定义 $\|\cdot\|_\beta\equiv(\|\cdot\|_\alpha)^D$ 为对偶范数. 设 $|||\cdot|||_\alpha$ 和 $|||\cdot|||_\beta$ 分别表示由 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 诱导的 M_n 上的矩阵范数. 利用(5.6.36)证明，$|||A^*|||_\beta=|||A|||_\alpha$ 对所有 $A\in M_n$ 成立. 证明 $|||A|||_2^2\leq|||A|||_\alpha|||A|||_\beta$ 对所有 $A\in M_n$ 成立，并且说明是怎样推广了习题 21 中的结果. 这个不等式与 $x=y$ 时的(5.4.13)有何关系？

23. 验证，下表中的各值给出的诸最佳常数 C_M 使得 $|||A|||_\alpha\leq C_M|||A|||_\beta$ 对所有 $A\in M_n$ 成立. 表中的所有范数都是矩阵范数. 表下面的提示 (i,j) 与表中确定的 (i,j) 元相配. 每种情形都给出了一个矩阵，对这个矩阵，不等式 $|||A|||_\alpha\leq C_M|||A|||_\beta$ 对常数 C_M 的给定值是等式.

[313]

$|||\cdot|||_\alpha$ \ $|||\cdot|||_\beta$	$|||\cdot|||_1$	$|||\cdot|||_2$	$|||\cdot|||_\infty$	$\|\cdot\|_1$	$\|\cdot\|_2$	$n\|\cdot\|_\infty$
$|||\cdot|||_1$	1	$\sqrt{n}$	n	1	$\sqrt{n}$	1
$|||\cdot|||_2$	$\sqrt{n}$	1	$\sqrt{n}$	1	1	1
$|||\cdot|||_\infty$	n	$\sqrt{n}$	1	1	$\sqrt{n}$	1
$\|\cdot\|_1$	n	$n^{3/2}$	n	1	n	n
$\|\cdot\|_2$	$\sqrt{n}$	$\sqrt{n}$	$\sqrt{n}$	1	1	1
$n\|\cdot\|_\infty$	n	n	n	n	n	1

下列矩阵都是 M_n 中的矩阵：

I 是单位矩阵；

J 是所有元素都是 1 的矩阵；

A_1 是其第 1 列的所有元素都是 1 而其余元素都是零的矩阵；

A_2 是其(1，1)元是 1，而其余元素都是零的矩阵.

(1，2) 根据(5.6.21)，可从(2，1)得到.

(1，3) $\|A\|_1 \leqslant \|A\|_1 \leqslant n\|A\|_1$； A_1.

(1，4) A_1.

(1，5) $\max\limits_{1\leqslant j\leqslant n}\left[\sum\limits_{i=1}^{n}|a_{ij}|\right]^2 \leqslant \sum\limits_{j=1}^{n}\left[\sum\limits_{i=1}^{n}|a_{ij}|\right]^2 \leqslant \left[\sum\limits_{j=1}^{n}1\right]\left[\sum\limits_{i=1}^{n}|a_{ij}|^2\right]$，
(Cauchy-Schwarz 不等式)； A_1.

(1，6) $\max\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^{n}|a_{ij}| \leqslant n\max\limits_{1\leqslant i,j\leqslant n}|a_{ij}|$； J.

(2，1) 从(2，5)和(5，1)可得； A_1^*.

(2，3) 从(2，5)和(5，3)可得； A_1.

(2，4) 从(2，5)和(5，4)可得； A_2.

(2，5) $\|A\|_2^2 = \rho(A^*A) \leqslant \sum\limits_{i=1}^{n}\lambda_i(A^*A) = \operatorname{tr} A^*A = \|A\|_2^2$； A_1.

(2，6) 从(2，5)和(5，6)可得； J.

(3，1) 根据(5.6.21)，从(1，3)可得： A_1^*.

(3，2) 根据(5.6.21)，从(2，3)可得； A_1^*.

(3，4) A_1^*.

(3，5) 类似于(1，5)； A_1^*.

(3，6) 类似于(1，6)； J.

[314]

(4，1) $\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{n}|a_{ij}| \leqslant n\max\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^{n}|a_{ij}|$： I.

(4，2) 从(4，5)和(5，2)可得；下述矩阵用两种方式给出等式：取 $a \equiv e^{2\pi i/n}$，注意$(\bar{a})^k = a^{-k}$，且当 $j\neq 0$ 时 $\sum\limits_{k=0}^{n-1}a^{kj} = 0$ 而 $j=0$ 时它等于 n；设 A 的$(k，j)$元是 a^{kj} 并且验证，$A^*A = nI$，$\|A\|_2 = \sqrt{n}$，$\|A\|_1 = n^2$ 和 $\|A\|_2 = n$.

(4，3) 类似于(4，1)： I.

(4，5) $\left[\sum\limits_{i,j=1}^{n}|a_{ij}|\right]^2 = \sum\limits_{i,j,p,q=1}^{n}|a_{ij}|\cdot|a_{pq}| \leqslant \frac{1}{2}\sum\limits_{i,j,p,q=1}^{n}\left[|a_{ij}|^2 + |a_{pq}|^2\right]$，
(算术-几何平均值不等式)； J.

(4，6) $\sum\limits_{i,j=1}^{n}|a_{ij}| \leqslant n^2\max\limits_{1\leqslant i,j\leqslant n}|a_{ij}|$； J.

(5, 1)　$\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2 \leqslant \sum_{j=1}^{n}\Big[\sum_{i=1}^{n}|a_{ij}|\Big]^2 \leqslant n\Big[\max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|a_{ij}|\Big]^2$;　I.

(5, 2)　$\sum_{i,j=1}^{n}|a_{ij}|^2 = \operatorname{tr} A^*A = \sum_{i=1}^{n}\lambda_i(A^*A) \leqslant n\lambda_{\max}(A^*A)$;　I.

(5, 3)　类似于(5, 1);　I.

(5, 4)　$\sum_{i,j=1}^{n}|a_{ij}|^2 \leqslant \Big[\sum_{i,j=1}^{n}|a_{ij}|\Big]^2$;　A_2.

(5, 6)　$\sum_{i,j=1}^{n}|a_{ij}|^2 \leqslant n^2\max_{1\leqslant i,j\leqslant n}|a_{ij}|^2$;　J.

(6, 1)　$\max_{1\leqslant i,j\leqslant n}|a_{ij}| \leqslant \max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|a_{ij}|$;　I.

(6, 2)　$\max_{1\leqslant i,j\leqslant n}|a_{ij}|^2 \leqslant \max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|a_{ij}|^2 = \max_{1\leqslant j\leqslant n}(A^*A)_{jj} \leqslant \rho(A^*A)$;　I.

(6, 3)　类似于(6, 1);　I.

(6, 4)　$\max_{1\leqslant i,j\leqslant n}|a_{ij}| \leqslant \sum_{i,j=1}^{n}|a_{ij}|$;　A_2.

(6, 5)　$\max_{1\leqslant i,j\leqslant n}|a_{ij}|^2 \leqslant \sum_{i,j=1}^{n}|a_{ij}|^2$;　A_2.

24. 证明习题23中界(5, 2)可改进为 $\|A\|_2 \leqslant [\operatorname{rank} A]^{1/2}|||A|||_2$. **提示**: $\operatorname{rank} A$ = A^*A 的非零特征值的个数.

315

25. 设 $A\in M_n$ 是给定的矩阵. 由引理(5.6.10)可知, 存在某个矩阵范数 $|||\cdot|||$ 使得 $\rho(A) \leqslant |||A||| \leqslant \rho(A)+\varepsilon$. 证明存在非奇异矩阵 $C=C(\varepsilon)\in M_n$ 使得 $\rho(A) \leqslant |||CAC^{-1}|||_2 \leqslant \rho(A)+\varepsilon$. **提示**: 利用与引理(5.6.10)中相同的构造法, 并且证明, 当 $\varepsilon\to 0$ 时 $|||CAC^{-1}|||_2^2 = \rho(A^*A)+O(\varepsilon)$.

26. 证明, $\|A\|_2^2 \geqslant \sum_{i=1}^{n}|\lambda_i|^2$ 对所有 $A\in M_n$ 成立, 其中等式成立, 当且仅当 A 是正规矩阵. 因为这个理由, 数值

$$\Big[\|A\|_2^2 - \sum_{i=1}^{n}|\lambda_i|^2\Big]^{1/2}$$

有时称为正规性亏损值. **提示**: 利用 Schur 三角化定理和 Frobenius 范数是酉不变范数的事实.

27. 利用定理(5.6.9)和友矩阵可以给出实系数或复系数多项式零点的界. 任意一个次数至少是1的多项式 $f(z)$ 可以写成 $f(z)=Cz^k p(z)$ 的形式, 其中 C 是非零常数,

$$p(z) = z^n + a_{n-1}z^{n-1} + a_{n-2}z^{n-2} + \cdots + a_1 z + a_0 \tag{5.6.38}$$

且 $a_0\neq 0$. $p(z)=0$ 的根是 $f(z)=0$ 的非零根, 并且对于这些根, 我们可以给出各种界. (a)证明, 友矩阵

$$C(p) = \begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_1 & -a_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & & 1 & 0 \end{bmatrix} \tag{5.6.39}$$

的特征多项式正好是 $p(z)$，因而 $C(p)$ 的特征值与 $p(z)=0$ 的根相同. **提示**：计算 $\det[zI-C(p)]$ 时用其第 1 列的代数余子式和归纳法. (b)利用定理(5.6.9)证明，若 z 是 $p(z)=0$ 的根且 $\|\cdot\|$ 是 M_n 上的任一矩阵范数，则 $|z|\leqslant\|C(p)\|$. 如下，z 表示 $p(z)=0$ 的任一根. (c)利用 $\|\cdot\|_1$ 证明

$$|z|\leqslant\max\{|a_0|,1+|a_1|,\cdots,1+|a_{n-1}|\}$$
$$\leqslant 1+\max\{|a_0|,|a_1|,\cdots,|a_{n-1}|\}. \tag{5.6.40}$$

这个关于根的界称为 Cauchy 界. (d)利用 $\|\cdot\|_\infty$ 证明

316

$$|z|\leqslant\max\{1,|a_0|+|a_1|+\cdots+|a_{n-1}|\}\leqslant 1+|a_0|+|a_1|+\cdots+|a_{n-1}|. \tag{5.6.41}$$

这个界称为 Montel 界. 证明它比 Cauchy 界要粗糙. (e)利用 $\|\cdot\|_1$ 证明

$$|z|\leqslant(n-1)+|a_0|+|a_1|+\cdots+|a_{n-1}|,$$

对于所有 $n>2$，这个界比(d)中的界粗糙. (f)利用 $\|\cdot\|_2$ 证明

$$|z|\leqslant[n-1+|a_0|^2+|a_1|^2+\cdots+|a_{n-1}|^2]^{1/2},$$

这个界比 Carmichael 和 Mason 界(5.6.42)粗糙. (g)利用 $n\|\cdot\|_\infty$ 证明

$$|z|\leqslant n\max\{1,|a_0|,|a_1|,\cdots,|a_{n-1}|\},$$

它比(5.6.41)中的界粗糙.

28. 利用习题 27 中相同的记号，我们可以改进其中(f)项中的界. 把友矩阵写成 $C(p)=S+R$，其中

$$S=\begin{bmatrix}0&0&\cdots&0&0\\1&0&\cdots&0&0\\0&1&\ddots&0&0\\\vdots&0&\ddots&\ddots&\vdots\\\vdots&\vdots&\ddots&\ddots&\vdots\\0&0&\cdots&1&0\end{bmatrix},$$

而

$$R=\begin{bmatrix}-a_{n-1}&-a_{n-2}&\cdots&-a_1&-a_0\\0&0&\cdots&0&0\\\vdots&\vdots&&\vdots&\vdots\\0&0&\cdots&0&0\end{bmatrix},$$

然后证明 $S^*R=R^*S=0$. 证明 $\|S^*S\|_2=1$ 和 $\|R^*R\|_2=|a_0|^2+|a_1|^2+\cdots+|a_{n-1}|^2$. 证明

$$\|C(p)\|_2^2=\|C(p)^*C(p)\|_2=\|(S+R)^*(S+R)\|_2$$
$$=\|S^*S+R^*R\|_2\leqslant\|S^*S\|_2+\|R^*R\|_2,$$

由此推导出 Carmichael 和 Mason 界

$$|z|\leqslant[1+|a_0|^2+|a_1|^2+\cdots+|a_{n-1}|^2]^{1/2} \tag{5.6.42}$$

317

29. 把界(5.6.41)应用到多项式

$$q(z) \equiv (z-1)p(z)$$
$$= z^{n+1} + (a_{n-1}-1)z^n + (a_{n-2}-a_{n-1})z^{n-1} + \cdots + (a_0-a_1)z + a_0,$$

然后证明

$$|\tilde{z}| \leqslant \max\{1, |a_0| + |a_0-a_1| + \cdots + |a_{n-2}-a_{n-1}| + |a_{n-1}-1|\}.$$

证明这个表示式的第 2 项不少于 1，并且推导出 Montel 的另一个界：

$$|\tilde{z}| \leqslant |a_0| + |a_0-a_1| + \cdots + |a_{n-2}-a_{n-1}| + |a_{n-1}-1| \tag{5.6.43}$$

30. 试用 Montel 界(5.6.43)证明 Kakeya 定理：若 $f(z)=a_nz^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0$ 是一个给定的多项式，其系数 a_i 均为非负实数且在 $a_n \geqslant a_{n-1} \geqslant \cdots \geqslant a_1 \geqslant a_0$ 的意义下是单调的，则 $f(z)=0$ 的所有根都在单位圆盘内，即 $|\tilde{z}| \leqslant 1$.

31. 前面四个习题给出的都是关于 $p(z)=0$ 的根的绝对值的上界，其实也可以得到它们的下界. 证明，若 $p(z)$是由(5.6.38)给出的 $a_0 \neq 0$ 的多项式，则函数

$$q(z) = \frac{1}{a_0}z^n p\left(\frac{1}{z}\right) = z^n + \frac{a_1}{a_0}z^{n-1} + \frac{a_2}{a_0}z^{n-2} + \cdots + \frac{a_{n-1}}{a_0}z + \frac{1}{a_0}$$

是一个次数为 n 的多项式，而它的各个根正好是 $p(z)=0$ 的各个根的倒数. 试用 $q(z)=0$ 的根的各相应上界得到 $p(z)=0$ 的根 $\tilde{z}$ 的下述下界.

Cauchy：

$$|\tilde{z}| \geqslant \frac{|a_0|}{\max\{1, |a_0| + |a_{n-1}|, |a_0| + |a_{n-2}|, \cdots, |a_0| + |a_1|\}}$$
$$\geqslant \frac{|a_0|}{|a_0| + \max\{1, |a_{n-1}|, |a_{n-2}|, \cdots, |a_1|\}}.$$

Montel：

$$|\tilde{z}| \geqslant \frac{|a_0|}{\max\{|a_0|, 1 + |a_1| + |a_2| + \cdots + |a_{n-1}|\}},$$
$$\geqslant \frac{|a_0|}{1 + |a_0| + |a_1| + \cdots + |a_{n-1}|}$$

Carmichael 和 Mason：

$$|\tilde{z}| \geqslant \frac{|a_0|}{[1 + |a_0|^2 + |a_1|^2 + \cdots + |a_{n-1}|^2]^{1/2}}.$$

318

32. 当把习题 31 中的下界与习题 27～30 中的上界结合起来时，就有可能把 $p(z)$的根确定在一个环域$\{z: r_1 \leqslant |z| \leqslant r_2\}$内. 例如，考虑

$$f(z) = \frac{1}{n!}z^n + \frac{1}{(n-1)!}z^{n-1} + \cdots + \frac{1}{2}z^2 + z + 1,$$

它是指数函数 e^z 的幂级数的 n 次部分和. 证明 $f(z)=0$ 的所有根 $\tilde{z}$ 满足不等式

$$\frac{1}{2} \leqslant |\tilde{z}| \leqslant 1 + n!.$$

把 Kakeya 定理应用于 $z^nf(1/z)$，试证明所有这些根实际上满足不等式 $|\tilde{z}| \geqslant 1$.

33. 因为对任何非奇异矩阵 D 有 $\rho(A)=\rho(D^{-1}AD)$，把习题 27 中所采用的方法应用到 $D^{-1}C(p)D$ 可以得到(5.6.38)中的多项式 $p(z)$的根的其他界. 为了计算方便，我们选取 $D=$

diag(p_1, p_2, …, p_n)，且所有 $p_i>0$，然后把 Cauchy 界(5.6.40)推广到

$$\begin{aligned}|\tilde{z}|\leqslant\max\Big\{&|a_0|\frac{p_n}{p_1},|a_1|\frac{p_{n-1}}{p_1}+\frac{p_{n-1}}{p_n},|a_2|\frac{p_{n-2}}{p_1}+\frac{p_{n-2}}{p_{n-1}},\cdots,\\&|a_{n-2}|\frac{p_2}{p_1}+\frac{p_2}{p_3},|a_{n-1}|+\frac{p_1}{p_2}\Big\},\end{aligned}\tag{5.6.44}$$

它对任何正参数 p_1, p_2, …, p_n 都成立.

34. 如果(5.6.38)中的所有系数 a_k 都不为零，选取 $p_k\equiv p_1/|a_{n-k+1}|$, $k=2$, 3, …, n, 试从(5.6.44)推导出 $p(z)$的根 $\tilde{z}$ 的 Kojima 界：

$$|\tilde{z}|\leqslant\max\left\{|a_0|,2\left|\frac{a_1}{a_2}\right|,2\left|\frac{a_2}{a_3}\right|,\cdots,2\left|\frac{a_{n-1}}{a_n}\right|\right\}.\tag{5.6.45}$$

35. 现在对某个 $r>0$，选取 $p_k\equiv r^k$, $k=1$, 2, …, n，并且证明，对任意 $r>0$，(5.6.44)蕴涵界

$$\begin{aligned}|\tilde{z}|&\leqslant\max\{|a_0|r^{n-1},|a_1|r^{n-2}+r^{-1},|a_2|r^{n-3}+r^{-1},\cdots,\\&\qquad|a_{n-2}|r+r^{-1},|a_{n-1}|+r^{-1}\}\\&\leqslant\frac{1}{r}+\max_{0\leqslant k\leqslant n-1}\{|a_k|r^{n-k-1}\}.\end{aligned}\tag{5.6.46}$$

36. 如果 $A\in M_n$，证明 Hermite 矩阵

$$\hat{A}=\begin{bmatrix}0&A\\A^*&0\end{bmatrix}\in M_{2n}$$

319

与 A 有相同的谱范数($\|\cdot\|_2$). **提示**：我们知道，一般地有 $\|\hat{A}\|_2=\rho(\hat{A}^*\hat{A})^{1/2}$.

37. 如果 A, $B\in M_n$, A 是非奇异矩阵，B 是奇异矩阵，又如果$\|\cdot\|$是任意矩阵范数，证明$\|A-B\|\geqslant1/\|A^{-1}\|$. **提示**：$B=A-(A-B)=A[I-A^{-1}(A-B)]$是奇异矩阵，所以$\|A^{-1}(A-B)\|\geqslant1$. 在 M_n 中它的几何意义是什么？如何用一个奇异矩阵去充分逼近一个非奇异矩阵？关于这个问题的进一步讨论见(7.4.1).

进一步阅读 习题 23 的表中的各个界取自下述文章 B. J. Stone, "Best Possible Ratios of Certain Matrix Norms," *Numerische Math.* 4(1962), 114-116，它还给出了另外一些界和参考文献. 有关用矩阵范数确定(习题 27～35 中)多项式的根的其他参考文献以及进一步的讨论可参看 M. Fujii and F. Kubo, "Operator Norms as Bounds for Roots of Algebraic Equations," *Proc. Japan Acad.* 49(1973), 805-808. 有关确定[定理(5.6.18)中]诱导范数间的界的问题的更一般讨论可见下文：H. Schneider and W. G. Strang, "Comparison Theorems for Supremum Norms," *Numerische Math.* 4(1962), 15-20. 在[Wie]中讨论了极小矩阵范数.

5.7 关于矩阵的向量范数

关于向量范数的所有公理虽然对于矩阵"大小"这一有用概念是必需的，但是，对于某些重要的应用来说，其中的矩阵范数的次乘性公理(4)不是必不可少的. 例如，非常有用的极限(5.6.14)实际上对比向量范数更一般的函数类仍成立，而不只限于矩阵范数. 为此，在这里着

重讨论关于矩阵的向量范数，即向量空间 M_n 上的向量范数(它可能没有次乘性质). 这样的范数常常称为广义矩阵范数. 我们将用 $\|\cdot\|$ 或用 $G(\cdot)$ 表示 M_n 上一般的向量范数，并且从 M_n 上的某些向量范数的例子开始. 它们可能是，也可能不是矩阵范数.

例 1　如果 $G(\cdot)$ 是 M_n 上的向量范数，又如果 $T, S\in M_n$ 是非奇异矩阵，则

$$G_{T,S}(A)\equiv G(TAS), A\in M_n \tag{5.7.1}$$

是 M_n 上的向量范数. 即使 $G(\cdot)$ 是矩阵范数，$G_{T,S}(\cdot)$ 也不一定有次乘性质.

[320]

练习　证明(5.7.1)中的 $G_{T,S}(\cdot)$ 总是 M^n 上的向量范数.

练习　设 $S=T=\frac{1}{2}I$，设 $G(\cdot)=\|\cdot\|$，证明 $G_{T,S}(\cdot)$ 不是矩阵范数.

练习　证明，如果 $G(\cdot)$ 是矩阵范数，且 $T=S^{-1}$，则 $G_{T,S}(\cdot)$ 是矩阵范数.

例 2　两个同阶矩阵 $A=[a_{ij}]$ 和 $B=[b_{ij}]$ 的 Hadamard 乘积正好是它们的对应元素的乘积 $A\circ B=[a_{ij}b_{ij}]$. 如果 $H\in M_n$ 是没有零元素的已知矩阵，$G(\cdot)$ 是 M_n 上的任意向量范数，则

$$G_H(A)=G(H\circ A) \tag{5.7.2}$$

是 M_n 上的向量范数. 即使 $G(\cdot)$ 是矩阵范数，$G_H(\cdot)$ 也不一定有次乘性质.

练习　证明(5.7.2)中的 $G_H(\cdot)$ 总是向量范数.

练习　证明(5.7.2)中的 $G_H(\cdot)$ 是，或者不是矩阵范数，取决于对 H 的选择. 考虑矩阵范数 $G(\cdot)=\|\cdot\|_1$，及 Hadamard 乘子矩阵

$$H_1=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} \text{或 } H_2=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \tag{5.7.3}$$

你可能想考察

$$A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}, B=\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} \text{和 } AB. \tag{5.7.4}$$

注意 $G_{H_1}(C)\leqslant G_{H_2}(C)$ 对所有 $C\in M_2$ 成立.

例 3　函数

$$G\left(\begin{bmatrix}a & b\\ c & d\end{bmatrix}\right)=\frac{1}{2}[|a+d|+|a-d|+|b|+|c|] \tag{5.7.5}$$

是 M_2 上的向量范数.

练习　证明(5.7.5)中的 $G(\cdot)$ 是向量范数，但不是矩阵范数. 你可能想考察(5.7.4)中的矩阵.

例 4　如果 $A\in M_n$ 是给定的矩阵，则集合 $F(A)\equiv\{x^*Ax: x\in \mathbf{C}^n \text{ 且 } x^*x=1\}$ 称为 A 的值域或数值范围，而函数

$$r(A)=\max_{x^*x=1}|x^*Ax|=\max_{z\in F(A)}|z| \tag{5.7.6}$$

[321]

称为 A 的数值半径.

练习　证明数值半径 $r(A)$ 是 M_n 上的向量范数. 提示：较难验证的部分只有正定性公理(1a)；见(4.1)节，习题 6. 但是，数值半径不是矩阵范数. 见习题 10.

例 5　M_n 上的 l_1 向量范数是

$$\|A\| = \max_{1\le i,j\le n} |a_{ij}|. \tag{5.7.7}$$

在(5.6)节中已经知道，$\|\cdot\|_\infty$ 是 M_n 上的向量范数而不是矩阵范数，不过 $n\|\cdot\|_\infty$ 是矩阵范数.

上述几个例子足以说明，M_n 确实有许多向量范数不是矩阵范数. 此外，在这些范数中，有些同时具有从次乘性得来的某些矩阵范数性质，有些则不具有. 但是，从 M_n 上的每个向量范数等价于任一矩阵范数(在它们有相同的收敛序列的意义下)；事实上，一个更为一般的结果可直接由定理(5.4.4)得到.

5.7.8 定理 设 f 是 M_n 上的准范数，即 M_n 上正定、齐次和连续的实值函数，另外设 $\|\cdot\|$ 是 M_n 上给定的矩阵范数. 则存在有限正常数 C_m 和 C_M，使得

$$C_m\|A\| \leqslant f(A) \leqslant C_M\|A\| \tag{5.7.9}$$

对所有 $A\in M_n$ 成立. 特别是当 $f(\cdot)$ 是 M_n 上的向量范数时，这两个不等式成立.

如果把有关矩阵范数的结论推广到关于矩阵的向量范数或更一般地推广到关于矩阵的准向量范数. 不等式(5.7.9)是很有用的. 例如，极限(5.6.14)就可以在这个意义下推广.

5.7.10 推论 如果 f 是 M_n 上的准范数，则对所有 $A\in M_n$，$\lim_{k\to\infty}[f(A^k)]^{1/k}$ 存在，且对所有 $A\in M_n$，

$$\lim_{k\to\infty}[f(A^k)]^{1/k} = \rho(A).$$

特别是，$f(\cdot)$ 是 M_n 上的向量范数时，这个极限成立.

证明：设 $\|\cdot\|$ 是 M_n 上的矩阵范数，考虑不等式

$$C_m\|A^k\| \leqslant f(A^k) \leqslant C_M\|A^k\|,$$

它蕴涵

$$C_m^{1/k}\|A^k\|^{1/k} \leqslant [f(A^k)]^{1/k} \leqslant C_M^{1/k}\|A^k\|^{1/k},$$

其中 $k=1, 2, 3, \cdots$. 但是，当 $k\to\infty$ 时，$C_m^{1/k}\to 1$，$C_M^{1/k}\to 1$ 和 $\|A^k\|^{1/k}\to\rho(A)$，所以我们得知，$\lim_{k\to\infty}[f(A^k)]^{1/k}$ 存在且等于 $\rho(A)$. □

在另一种意义下，M_n 上的任一个向量范数都等价于一个矩阵范数. 这在前面的例 5 中已举例作了说明. 可以用常数因子 n 修改向量范数 $\|\cdot\|_\infty$ 使它成矩阵范数. 这种情况不是偶然的；每个向量范数都可以这样修改.

322

5.7.11 定理 对于 M_n 上每个向量范数 $G(\cdot)$，存在有限正常数 $c(G)$，使得 $c(G)G(\cdot)$ 是 M_n 上的矩阵范数. 如果 $\|\cdot\|$ 是 M_n 上的矩阵范数，又如果对所有 $A\in M_n$ 有

$$C_m\|A\| \leqslant G(A) \leqslant C_M\|A\|, \tag{5.7.11a}$$

则

$$c(G) \leqslant \frac{C_M}{C_m^2}.$$

此外，存在一个矩阵范数，使得关于 $c(G)$ 的这个上界是可以达到的，所以

$$c(G) = \min\left\{\frac{C_M}{C_m^2} : \|\cdot\| \text{是矩阵范数且(5.7.11a)成立}\right\}.$$

证明：对任意 $c>0$，函数$\|\cdot\|=cG(\cdot)$可能不满足次乘性公理，但它满足矩阵范数的所有其他公理. 此外，很容易从 $G(\cdot)$的连续性和 $G(\cdot)$的单位球的紧性推导出

$$c(G)\equiv\max_{A\neq 0\neq B}\frac{G(AB)}{G(A)G(B)}=\max_{G(A)=1=G(B)}G(AB)$$

是有限正数. 于是，对所有 $A,B\in M_n$，

$$G(AB)\leqslant c(G)G(A)G(B)\text{ 和 }c(G)G(AB)\leqslant c(G)G(A)c(G)G(B)$$

假定$\|\cdot\|$是 M_n 上的矩阵范数，且假定给出了 $G(\cdot)$与$\|\cdot\|$间的不等式. 于是

$$G(AB)\leqslant C_M\|AB\|\leqslant C_M\|A\|\|B\|\leqslant\frac{C_M}{C_m^2}G(A)G(B),$$

323 因而

$$c(G)\leqslant\frac{C_M}{C_m^2}.$$

如果我们特别选取矩范阵数$\|\cdot\|=c(G)G(\cdot)$，则 $C_M=c(G)$，且 $C_m=1/c(G)$，所以 $C_M/C_m^2=c(G)$. □

练习 证明，如果 $k\geqslant c(G)$，则 $kG(\cdot)$是矩阵范数. 特别地，证明 $C_MG(\cdot)/C_m^2$ 总是矩阵范数.

练习 试直接由(5.7.11)推导出(5.7.10)中的有关向量范数的结果.

矩阵范数次乘性的一个推论是下述事实：相应于 M_n 上的每种矩阵范数，都有 $\mathbf{C}^n$ 上的某个相容向量范数. 次乘性的另一个推论是，对每个矩阵范数$\|\cdot\|$有$\|A\|\geqslant\rho(A)$；M_n 上的向量范数 $G(\cdot)$如果对所有 $A\in M_n$ 有$\|A\|\geqslant\rho(A)$，就称它为谱优势的. 有趣的是 M_n 上的有些向量范数在 $\mathbf{C}^n$ 上有相容向量范数，而有些则没有. 那些在 $\mathbf{C}^n$ 上没有相容的向量范数的广义范数当中，有些是谱优势的，而有些则不是. 于是，M_n 上的向量范数在 $\mathbf{C}^n$ 上可能有相容的向量范数，而它又不是矩阵范数.

5.7.12 定义 $\mathbf{C}^n$ 上的向量范数$\|\cdot\|$称为与 M_n 上的向量范数 $G(\cdot)$是相容的，指的是对所有 $x\in\mathbf{C}^n$ 和所有 $A\in M_n$ 有

$$\|Ax\|\leqslant G(A)\|x\|.$$

有时采用术语协和的，而有时称向量范数$\|\cdot\|$从属于广义矩阵范数 $G(\cdot)$. 在前一节[例如，(5.6.27)，(5.6.30)]中，已经述及了这些概念. 在这里要特别论述几个有关的结果.

5.7.13 定理 如果$\|\cdot\|$是 M_n 上的矩阵范数，则在 $\mathbf{C}^n$ 上存在某个与它相容的向量范数.

证明：如果我们定义$|x|\equiv\|[x\ 0\ 0\ \cdots\ 0]\|$，那么$|Ax|=\|[Ax\ 0\ \cdots\ 0]\|=\|A[x\ 0\ \cdots\ 0]\|\leqslant\|A\|\|[x\ 0\ \cdots\ 0]\|=\|A\||x|$. □

我们已经知道了它的逆命题. 定理(5.6.2)说明，如果$\|\cdot\|$是 $\mathbf{C}^n$ 上给定的向量范数，则存在与它相容的矩阵范数[诱导范数(5.6.1)].

练习 试说明由(5.7.13)确保的 $\mathbf{C}^n$ 上的相容向量范数不一定唯一. 实际上，$|x|\equiv$ 324 $\|[x\ x\ \cdots\ x]\|$也是可行的，因为对任意非零向量 $y\in\mathbf{C}^n$ 有$|x|\equiv\|xy^*\|$.

5.7.14 定理　设 $G(\cdot)$ 是 M_n 上的向量范数，且在 $\mathbf{C}^n$ 上有相容向量范数 $\|\cdot\|$．则 $G(A)\geqslant\rho(A)$ 对所有 $A\in M_n$ 成立．更一般地，

$$G(A_1)G(A_2)\cdots G(A_k)\geqslant\rho(A_1A_2\cdots A_k)\tag{5.7.15}$$

对所有 $A_1, A_2, \cdots, A_k\in M_n$ 和所有 $k=1, 2, \cdots$ 成立．

证明： 假定 $k=2$，且设 $x\in\mathbf{C}^n$ 是适合 $A_1A_2x=\lambda x$ 的非零向量，其中 $|\lambda|=\rho(A_1A_2)$．则

$$\begin{aligned}\rho(A_1A_2)\|x\|&=\|\lambda x\|=\|A_1A_2x\|\\&=\|A_1(A_2x)\|\leqslant G(A_1)\|A_2x\|\leqslant G(A_1)G(A_2)\|x\|.\end{aligned}$$

因为 $\|x\|\neq0$，由此得出 $\rho(A_1A_2)\leqslant G(A_1)G(A_2)$．类似地用归纳法可以证明一般情形．　□

什么时候 M_n 上给定的向量范数在 $\mathbf{C}^n$ 上有相容的向量范数呢？条件(5.7.15)是必要的：为了证明它也是充分条件，需要一个技巧性引理．

5.7.16 引理　设 $G(\cdot)$ 是 M_n 上适合(5.7.15)的向量范数，又设 $\|\cdot\|_2$ 表示 M_n 上的谱范数．则存在有限正常数 $c=c(G)$，使得

$$G(A_1)G(A_2)\cdots G(A_k)\geqslant c\|A_1A_2\cdots A_k\|_2$$

对所有 $A_1, A_2, \cdots, A_k\in M_n$ 和所有 $k=1, 2, \cdots$ 成立．

证明： 根据推论(5.4.5)，存在有限正常数 $b=b(G)$，使得 $\|A\|_2\geqslant bG(A)$ 对所有 $A\in M_n$ 成立．设 k 是给定的正整数，且设 $A_1, A_2, \cdots, A_k\in M_n$ 是给定的矩阵．根据奇异值分解定理(7.3.5)，有酉矩阵 V，W 以及对角矩阵 $\Sigma=\mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$，其中，所有 $\sigma_i\geqslant0$，使得 $A_1A_2\cdots A_k=V\Sigma W^*$ 和 $\rho(\Sigma)=\max\{\sigma_1, \sigma_2, \cdots, \sigma_n\}=\|A_1A_2\cdots A_k\|_2$．根据(5.7.15)，有

$$\begin{aligned}G(V^*)G(A_1)G(A_2)\cdots G(A_k)G(W)&\geqslant\rho(V^*A_1A_2\cdots A_kW)\\&=\rho(\Sigma)\\&=\|\Sigma\|_2\\&=\|V^*A_1A_2\cdots A_kW\|_2\\&=\|A_1A_2\cdots A_k\|_2.\end{aligned}$$

后一个等式成立是因为谱范数是酉不变的．因而得出，

$$\begin{aligned}G(A_1)G(A_2)\cdots G(A_k)&\geqslant\frac{1}{G(V^*)G(W)}\|A_1A_2\cdots A_k\|_2\geqslant\frac{b^2}{\|V^*\|_2\|W\|_2}\|A_1A_2\cdots A_k\|_2\\&=b^2\|A_1A_2\cdots A_k\|_2.\end{aligned}$$

如果取 $c\equiv b^2$，引理的结论就证明了．　□

325

5.7.17 定理　设 $G(\cdot)$ 是 M_n 上的向量范数．则存在 $\mathbf{C}^n$ 上的向量范数 $\|\cdot\|$ 使得

$$\|Ax\|\leqslant G(A)\|x\|$$

对所有 $x\in\mathbf{C}^n$ 和所有 $A\in M_n$ 成立，当且仅当

$$G(A_1)G(A_2)\cdots G(A_k)\geqslant\rho(A_1A_2\cdots A_k)$$

对所有 $A_1, A_2, \cdots, A_k\in M_n$ 和所有 $k=1, 2, \cdots$ 成立．

证明： 必要性已在定理(5.7.4)中证明．关于充分性，需要证明，存在 M_n 上的矩阵范数 $\|\cdot\|$ 使得 $G(A)\geqslant\|A\|$ 对所有 $A\in M_n$ 成立．设 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上与 $\|\cdot\|$ 相容的向量范数[定理

(5.7.13)保证它存在]，且设 $x\in\mathbf{C}^n$ 和 $A\in M_n$ 是给定的. 因此，如果能构造出以 $G(\cdot)$ 为优势范数的矩阵范数，就证明了 $\|Ax\|\leqslant\|A\|\|x\|\leqslant G(A)\|x\|$.

对于给定的矩阵 $A\in M_n$，有无数多种方式把 A 表示成若干个矩阵的乘积或若干个矩阵的乘积之和. 定义

$$\|A\|\equiv\inf\Big\{\sum_i G(A_{i1})\cdots G(A_{ik_i}):\sum_i A_{i1}\cdots A_{ik_i}=A,\text{且所有 }A_{ik_j}\in M_n\Big\}.$$

如果 $\sum A_{i1}\cdots A_{ik_i}=A$，则根据引理(5.7.16)和关于谱范数的三角不等式，有

$$\sum_i G(A_{i1})\cdots G(A_{ik_i})\geqslant\sum_i c\,\|A_{i1}\cdots A_{ik_i}\|_2\geqslant c\,\Big\|\sum_i A_{i1}\cdots A_{ik_i}\Big\|_2=c\|A\|_2.$$

由这个不等式可推出，所构造的函数 $\|\cdot\|$ 是正定的. $\|\cdot\|$ 的齐次性可直接从 $G(\cdot)$ 的齐次性得出. 关于 $\|\cdot\|$ 的三角不等式和次乘性可由它作为乘积之和的下确界的定义得出. □

326

练习 详细证明，上述定理中所构造的函数 $\|\cdot\|$ 满足三角不等式，且有次乘性质. **提示:** 如果 $C=A+B$ 或 $C=AB$，则(分别)作为乘积之和的 A 与 B 的每个表示式给出 C 作为乘积之和的表示式，但是，并非 C 的所有这些表示式都以这种形式出现.

练习 考虑 M_2 上的向量范数(5.7.5)，假如它在 $\mathbf{C}^2$ 上有相容向量范数 $\|\cdot\|$，则可证明

$$\left\|\begin{bmatrix}1\\0\end{bmatrix}\right\|=\left\|\begin{bmatrix}0&1\\0&0\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}\right\|\leqslant G\begin{bmatrix}0&1\\0&0\end{bmatrix}\left\|\begin{bmatrix}0\\1\end{bmatrix}\right\|,$$

且

$$\left\|\begin{bmatrix}0\\1\end{bmatrix}\right\|=\left\|\begin{bmatrix}0&0\\1&0\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\right\|\leqslant G\begin{bmatrix}0&0\\1&0\end{bmatrix}\left\|\begin{bmatrix}1\\0\end{bmatrix}\right\|,$$

它蕴涵

$$\left\|\begin{bmatrix}1\\0\end{bmatrix}\right\|\leqslant G\begin{bmatrix}0&1\\0&0\end{bmatrix}G\begin{bmatrix}0&0\\1&0\end{bmatrix}\left\|\begin{bmatrix}1\\0\end{bmatrix}\right\|,$$

因而

$$1\leqslant G\begin{bmatrix}0&1\\0&0\end{bmatrix}G\begin{bmatrix}0&0\\1&0\end{bmatrix}.$$

试证明这是不能成立的，由此可知 M_2 上的向量范数 $G(\cdot)$ 在 $\mathbf{C}^2$ 上不可能有相容向量范数.

练习(续) 即使 M_2 上的向量范数(5.7.5)在 $\mathbf{C}^2$ 上没有任何相容向量范数，也可证明它是谱优势的. 试根据定理(5.7.17)进行讨论.

现在，知道了关于 M_n 上的向量范数在 $\mathbf{C}^n$ 上有相容向量范数的有用的必要充分条件. 同时也知道，只要在 $\mathbf{C}^n$ 上有一个向量范数，则诱导矩阵范数(5.6.1)就是与它相容的 M_n 上的次乘性向量范数. 那么，什么时候 $\mathbf{C}^n$ 上的向量范数在 M_n 上有非次乘性的相容向量范数呢？这样的范数总是有的.

5.7.18 定理 设 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上给定的向量范数. 则在 M_n 上存在不是矩阵范数的向量范数 $G(\cdot)$，使得

327

$$\|Ax\|\leqslant G(A)\|x\|$$

对所有的 $x\in\mathbf{C}^n$ 和 $A\in M_n$ 成立.

证明：设 $P\in M_n$ 是具有零主对角线的任一置换矩阵，例如，$P=[p_{ij}]$，其中，如果 $j=i+1$，或 $i=n$ 且 $j=1$，则 $p_{ij}=1$；否则，$p_{ij}=0$. 设 $\|\cdot\|$ 表示 M_n 上由向量范数 $\|\cdot\|$ [通过(5.6.1)]诱导的矩阵范数. 在 M_n 上定义 $G(\cdot)$ 为

$$G(A)\equiv\|A\|+\|P\|\|P^T\|\max_{1\le i\le n}|a_{ii}|.$$

显然，$G(\cdot)$ 是 M_n 上的向量范数，$G(A)\geq\|A\|$ 对所有 $A\in M_n$ 成立，且

$$\|Ax\|\leqslant\|A\|\|x\|\leqslant G(A)\|x\|$$

对所有 $A\in M_n$ 和所有 $x\in\mathbf{C}^n$ 成立. 但是，

$$G(PP^T)=G(I)=\|I\|+\|P\|\|P^T\|=1+\|P\|\|P^T\|$$

$$G(P)=\|P\|$$

$$G(P^T)=\|P^T\|$$

$$G(PP^T)>G(P)G(P^T).$$

因此，M_n 上的范数 $G(\cdot)$ 与 $\mathbf{C}^n$ 上给定的向量范数 $\|\cdot\|$ 相容，而它不是次乘性范数. □

练习 设 $A=[a_{ij}]\in M_n$，考虑极大行和矩阵范数的下述变更：

$$\|A\|\equiv\|A+\operatorname{diag}(a_{11},a_{22},\cdots a_{nn})\|_\infty.$$

证明这是形如(5.7.2)的 Hadamard 乘积范数，因而它是 M_n 上的向量范数. 证明这个范数与 $\mathbf{C}^n$ 上的向量范数 $\|\cdot\|_\infty$ 相容. 然后计算

$$\left\|\begin{bmatrix}0&1\\1&0\end{bmatrix}\right\| \text{ 和 } \left\|\begin{bmatrix}0&1\\1&0\end{bmatrix}^2\right\|$$

并且证明这个范数不是次乘性范数.

习题

1. 设 $G(\cdot)$ 是 M_n 上的向量范数，设 $y\in\mathbf{C}^n$ 是给定的非零向量. 证明函数

$$\|x\|\equiv G(xy^*)$$

是 $\mathbf{C}^n$ 上的向量范数. 当

$$y=[1,1,\cdots,1]^T \text{ 或 } y=[1,0,0,\cdots,0]^T$$

时，这个范数是什么？

[328]

2. 设 $\|\cdot\|$ 是 M_n 上的任意向量范数，设 $A\in M_n$ 是给定的矩阵，且 $\varepsilon>0$ 是给定的数. 证明存在某个 $K=K(\varepsilon, A)>0$，使得

$$[\rho(A)-\varepsilon]^k\leqslant\|A^k\|\leqslant[\rho(A)+\varepsilon]^k$$

对所有 $k>K$ 成立.

3. 设 $\|\cdot\|$ 是 M_n 上的任意向量范数，且设 $A\in M_n$ 是给定的矩阵. (a)利用习题 2 证明，如果 $\rho(A)<1$，则当 $k\to\infty$ 时 $\|A^k\|\to0$. 它以什么速度收敛？(b)反过来，如果当 $k\to\infty$ 时 $\|A^k\|\to0$，证明 $\rho(A)<1$. **提示**：如果 $Ax=\lambda x$ 且 $x\neq0$，考虑 $\|A^k[x\ \cdots\ x]\|$. (c)关于利用向量范数讨论矩阵幂级数的收敛性，你能说些什么？

4. 设 $G(\cdot)$ 是 M_n 上给定的向量范数，且定义函数 G'：$M_n\to\mathbf{R}$ 为

$$G'(B) \equiv \max_{G(A)=1} G(BA).$$

证明 $G'(\cdot)$ 总是 M_n 上的矩阵范数．证明总有 $G'(I)=1$．如果 $G(I)=1$，证明 $G'(B)\geqslant G(B)$ 对所有 $B\in M_n$ 成立．

以下四个习题是习题4的延伸．

5．如果 $G(\cdot)$ 是 M_n 上的矩阵范数，证明 $G'(B)\leqslant G(B)$ 对所有 $B\in M_n$ 成立，并且证明，如果 $G(I)=1$，则 $G'(\cdot)=G(\cdot)$．

6．证明总有 $G''(\cdot)=G'(\cdot)$．

7．如果 $G(\cdot)$ 是 M_n 上的向量范数，且 $G(I)=1$，证明 $G(\cdot)$ 是矩阵范数，当且仅当 $G'(B)\leqslant G(B)$ 对所有 $B\in M_n$ 成立．

8．证明，在习题4中定义 $G'(\cdot)$ 时，可以颠倒 A 与 B 出现的顺序，从而得到另一个矩阵范数．用例子说明，这另一个范数不一定等于 $G'(\cdot)$．

9．证明与 M_n 上一个给定的向量范数相容的 $\mathbf{C}^n$ 上的所有向量半范数组成的集合是凸集—实际上它是一个锥．

10．说明数值半径 $r(\cdot)$ 不是 M_n 上的矩阵范数．试考察(5.7.4)中的各矩阵且比较 $r(AB)$ 与 $r(A)r(B)$．

11．定理(5.6.9)中的不等式 $\|A\|\geqslant\rho(A)$ 可以由关于矩阵范数 $\|\cdot\|$ 的次乘性公理(4)得出．但是，可能有这种情形，M_n 上的一个向量范数满足这个不等式(即它是谱优势的)，而它不是矩阵范数．证明 $r(A)\geqslant\rho(A)$ 对每个 $A\in M_n$ 成立．更一般地，证明 $\sigma(A)\subset\{x^*Ax: x^*x=1\}$．

12．证明 M_n 上的向量范数 $\|\cdot\|_\infty$ 在 $\mathbf{C}^n$ 上不可能有任何相容的量范数．**提示**：考虑 $\|J_n\|_\infty$ 和 $\rho(J_n)$．但是，证明 $n\|\cdot\|_\infty$ 是 M_n 上的矩阵范数，因而它在 $\mathbf{C}^n$ 上有相容向量范数．

329

13．设 $A=[a_{ij}]\in M_{m,n}$，A 的第 i 行的转置用 $r_i(A)=[a_{i1}, a_{i2}, \cdots, a_{in}]^T$ 表示，而 A 的第 j 列用 $c_j(A)=[a_{1j}, a_{2j}, \cdots, a_{mj}]^T$ 表示．又假定 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 分别是 $\mathbf{C}^n$ 和 $\mathbf{C}^m$ 上的向量范数，然后，定义 $G_{\beta,\alpha}: M_{m,n}\to\mathbf{R}$ 为

$$G_{\beta,\alpha}(A) \equiv \|[\|r_1(A)\|_\alpha, \|r_2(A)\|_\alpha, \cdots, \|r_m(A)\|_\alpha]^T\|_\beta.$$

类似地，定义 $G^{\alpha,\beta}: M_{m,n}\to\mathbf{R}$ 为

$$G^{\alpha,\beta}(A) \equiv \|[\|c_1(A)\|_\beta, \|c_2(A)\|_\beta, \cdots, \|c_n(A)\|_\beta]^T\|_\alpha.$$

证明，$G_{\beta,\alpha}(\cdot)$ 和 $G^{\alpha,\beta}(\cdot)$ 各为 $M_{m,n}$ 上的向量范数，但 $G^{\alpha,\beta}(\cdot)$ 与 $G_{\alpha,\beta}(\cdot)$ 不一定相同．注意这是在矩阵空间上定义向量范数的自然方式．

14．试比较习题13中的 $G_{\beta,\alpha}(\cdot)$ 与(5.6)节习题4中定义的范数 $\|\cdot\|_{\alpha,\beta}$，并且用例子说明，即使 $m=n$ (且 $\|\cdot\|_\alpha=\|\cdot\|_\beta$)，$G_{\beta,\alpha}(\cdot)$ 也不一定是 M_n 上的矩阵范数．

15．在习题13中，当 $\|\cdot\|_\alpha=\|\cdot\|_2=\|\cdot\|_\beta$ 时，$G_{\beta,\alpha}(\cdot)$ 是什么范数？$G^{\alpha,\beta}(\cdot)$ 是什么范数？

16．在习题13中，当 $\|\cdot\|_\alpha=\|\cdot\|_1$ 且 $\|\cdot\|_\beta=\|\cdot\|_\infty$ 时，$G_{\beta,\alpha}(\cdot)$ 是什么范数？$G^{\beta,\alpha}(\cdot)$ 是什么范数？$G_{\alpha,\beta}(\cdot)$ 与 $G^{\alpha,\beta}(\cdot)$ 有什么关系？

17．若 $G(\cdot)$ 是 M_n 上的向量范数，定义 $G(\cdot)$ 的谱示性数为

$$m(G) \equiv \max_{G(A)\leqslant 1} \rho(A)$$

证明，$G(\cdot)$是谱优势的当且仅当$m(G)\leqslant 1$. 然后证明，M_n上的任一向量范数在乘上一个常数[其中最小常数一定是$m(G)$]后可以变成一个谱优势范数. 如果$m(G)=1$，则称M_n上的范数$G(\cdot)$为极小谱优势的.

18. 证明，任何诱导矩阵范数是习题17中所定义的极小谱优势范数. 说明有一些范数是极小谱优势范数但不是诱导范数. 证明数值半径$r(A)$是极小谱优势的.

19. 证明谱示性数是由M_n上的向量范数组成的锥上的凸函数，因而证明M_n上的所有谱优势向量范数组成的集合是凸集.

330

20. 证明，M_n上的向量范数$G(\cdot)$是谱优势的，当且仅当对每个$A\in M_n$存在一个常数r_A[只与$G(\cdot)$和A有关]，使得对所有整数$k>0$，有

$$G(A^k)\leqslant r_A G(A)^k.$$

21. (a)证明，当A是正规矩阵时，数值半径$r(\cdot)$适合$r(A)=\rho(A)=\|A\|_2$，但是一般$r(A)\leqslant\|A\|_2$. 给出一个关于矩阵$A\in M_n$的例子使得$r(A)<\|A\|_2$. **提示**：证明当$U\in M_n$是酉矩阵时$r(U^*AU)=r(A)$，然后利用A可酉对角化的事实. 再说明一般地有

$$r(A)=\max_{\|x\|_2=1}|x^*Ax|\leqslant\max_{\|x\|_2=1}\|Ax\|_2\|x\|_2=\|A\|_2.$$

(b)证明，对所有$A\in M_n$都有$r(A)=r(A^*)$. (c)按如下所述证明对所有$A\in M_n$都有$\|A\|_2\leqslant 2r(A)$：记

$$A=(A+A^*)/2+(A-A^*)/2\equiv A_1+A_2,$$

并注意到A_1和A_2是正规矩阵. 然后证明

$$\|A\|_2\leqslant\|A_1\|_2+\|A_2\|_2=r(A_1)+r(A_2)\leqslant r(A)+r(A^*)=2r(A).$$

(d)试考察适当的$n\times n$矩阵，如$\begin{bmatrix}1&0\\0&0\end{bmatrix}$和$\begin{bmatrix}0&1\\0&0\end{bmatrix}$，说明(a)和(c)中给出的界

$$\frac{1}{2}\|A\|_2\leqslant r(A)\leqslant\|A\|_2 \tag{i}$$

是可以达到的.

22. 试用习题21(d)中的不等式以及定理(5.7.11)中给出的关于$c(r)$的界证明，函数$4r(\cdot)$是M_n上的矩阵范数. 试考察$A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$，$A^*$和$AA^*$，证明$c(r)=4$.

23. 由习题21(d)中的(i)以及(5.6)节习题23给出的不等式，

$$\frac{1}{\sqrt{n}}\|A\|_2\leqslant\|A\|_2\leqslant\|A\|_2 \tag{ii}$$

试推出，对所有的$A\in M_n$，不等式

$$\frac{1}{2\sqrt{n}}\|A\|_2\leqslant r(A)\leqslant\|A\|_2 \tag{iii}$$

成立，并且证明其上界是可以达到的. 验证$A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$和$A=I$分别给出了(i)和(ii)的下界中

等式成立的例子，然后验证 $A=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$ 是(i)和(ii)的上界中等式成立的例子. 由此说明，为什么(iii)中的上界是可以达到的，而(iii)中的下界不一定能达到. 为什么对所有的 $A\in M_n$ 存在一个有限正常数 c_n 使得 $c_n\|A\|_2\leqslant r(A)$? 实际上，n 为偶数时 $c_n=(2n)^{-1/2}$，而 n 为奇数时 $c_n=(2n-1)^{-1/2}$. 对于偶数 n，相等的情形要求矩阵酉相似于形如 $r(A)\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$ 的诸矩阵的直和；当 n 为奇数时，矩阵的直和一定要附加一个 1×1 的被加单项 $[\alpha]$，且 $|\alpha|=r(A)$.

331

24. 证明 $[A,B]\equiv \operatorname{tr} AB^*$ 定义 M_n 上的一个内积，而 M_n 上的 l_2 范数可由 $[\cdot,\cdot]$ 导出；即 $\|A\|_2=[A,A]^{1/2}$ 对所有 $A\in M_n$ 成立. 证明，如果 $X=xx^*$ 是秩 1 Hermite 矩阵，则 $\|X\|_2=\|x\|_2^2$. 证明给定的矩阵 $A\in M_n$ 的值域正好是 A 到范数为 1 的秩 1 Hermite 矩阵的集合(关于内积 $[\cdot,\cdot]$)的射影组成的集合，并且证明 $r(A)=\max\{|[A,X]|: X$ 是秩 1 Hermite 矩阵且 $\|X\|_2=1\}$. 利用 Cauchy-Schwarz 不等式证明 $r(A)\leqslant\|A\|_2$.

25. 数值半径与自然逼近问题有联系. 设 $A\in M_n$ 是给定的，并且假定我们想在最小二乘方的意义下用一个秩为 1 的 Hermite 矩阵的纯量倍去尽可能地逼近 A. 如果记 $X=cxx^*$, $c\in\mathbf{C}$, $\|x\|_2=1$，试证明，当 $c=[A,\tilde{x}\tilde{x}^*]$ 且其中的单位向量 $\tilde{x}$ 可使(5.7.6)中的极大值达到时，则

$$\|A-X\|_2^2=\|A-cxx^*\|_2^2\geqslant\|A\|_2^2-2|c[A,xx^*]|+|c|^2$$

可以达到极小. 由此断定，如果对所有纯量 c 以及所有具有 $\|X\|_2=1$ 的秩 1 Hermite 矩阵，$\|A-cX\|_2$ 可以达到极小，则 $|c|=r(A)$.

26. 前面两个习题引出了数值半径和值域的一个自然推广. 设 $\Phi\subset M_n$ 是一个非空的矩阵集合，且适合以下条件：

(a) 若 $X\in\Phi$，则对所有 $a\in\mathbf{C}$ 有 $aX\in\Phi$;

(b) 对所有 $X\in\Phi$ 有 $[A,X]=\operatorname{tr} AX^*=0$ 当且仅当 $A=0$;

(c) Φ 是一个闭集.

若 $A\in M_n$，定义

$$\phi(A)\equiv\max_{\substack{X\in\Phi\\ \|X\|_2\leq 1}}|[A,X]|=\max_{\substack{X\in\Phi\\ \|X\|_2\leq 1}}|\operatorname{tr} AX^*|.$$

说明 $\phi(\cdot)$ 是有意义的，它是 M_n 上的向量范数，且满足 $|\phi(A)|\leqslant\|A\|_2$. 证明对每个 $A\in M_n$ 都存在某个 $X_A\in\Phi$ 使得 $\|X_A\|_2=1$ 且 $\phi(A)=|[A,X_A]|$.

332

考虑用 Φ 中的矩阵去逼近给定的矩阵 $A\in M_n$，即求某个 $X\in\Phi$，使得 $\|A-X\|_2$ 达到极小. 证明，矩阵 $\phi(A)X_A$ 给出了一个最佳逼近，其中 $\phi(A)=|[A,X_A]|$，并且用任意 $X\in\Phi$ 逼近 A 的误差有最大下界 $\|A-X\|_2^2\geqslant\|A\|_2^2-|[A,X_A]|^2\geqslant 0$.

证明，若 Φ 是由秩 1 Hermite 矩阵的所有纯量倍组成的集合，则 $\phi(A)=r(A)$. 在 Φ 是酉矩阵的所有纯量倍的集合的情形将在例(7.4.6)中讨论；在这种情形下 $\phi(A)$ 是 A 的各奇异值的平均值. 另一个值得注意的情形是 Φ 是所有奇异矩阵的集合，这将在例(7.4.1)中讨论，在这种情形下 $\phi(A)$ 是 A 的最小奇异值. 还有一个值得注意的情形是，秩是已知的正定矩阵、Hermite 矩阵或正规矩阵的所有纯量倍的集合，或者是由酉相似于一个给定的矩阵的所有矩阵

的所有纯量倍组成的集合．在上述每一种情形，与值域类似的是集合$\{[A, X]: X\in\Phi\}$.

27. 即使数值半径$r(A)$不是矩阵范数，但对所有$m=1, 2, \cdots$和所有$A\in M_n$，它的确满足幂不等式$r(A^m)\leqslant[r(A)]^m$，试按以下步骤证明这个结论．

(a) 说明只要证明下述结论就足够了：若$r(A)\leqslant 1$，则对所有$m=1, 2, \cdots$有$r(A^m)\leqslant 1$.

设$m\geqslant 2$是给定的正整数，转到余下的论证，设$\{w_k\}=\{e^{2\pi i k/m}\}_{k=1}^m$表示$m$次单位根的集合，注意到$\{w_k\}$是一个有限乘法群，并且对每个$j=1, 2, \cdots, m$有$\{w_j w_k\}_{k=1}^m=\{w_k\}_{k=1}^m$.

(b) 说明

$$1-z^m=\prod_{k=1}^m(1-w_k z),$$

然后证明，对所有$z\in\mathbf{C}$有

$$p(z)\equiv\frac{1}{m}\sum_{j=1}^m\prod_{\substack{k=1\\k\neq j}}^m(1-w_k z)\equiv 1.$$

提示：注意到$p(z)$是次数至多为$m-1$的多项式，且

$$p(z)=\frac{1}{m}\sum_{j=1}^m\frac{1-z^m}{1-w_j z}$$

对所有$z\in\mathbf{C}$都有$p(z)=p(w_1 z)=\cdots=p(w_m z)$，因而$p(z)=$常数$=p(0)=1$.

(c) 证明

$$I-A^m=\prod_{k=1}^m(I-w_k A)\quad 和\quad I=\frac{1}{m}\sum_{j=1}^m\prod_{\substack{k=1\\k\neq j}}^m(I-w_k A).$$

(d) 设$x\in\mathbf{C}^n$是任一单位向量，$\|x\|_2=1$，又设$A\in M_n$. 验证

$$\begin{aligned}1-x^*A^m x&=x^*(I-A^m)x=(Ix)^*(I-A^m)x\\&=\left[\frac{1}{m}\sum_{j=1}^m\prod_{\substack{k=1\\k\neq j}}^m(I-w_k A)x\right]^*\left[\prod_{k=1}^m(I-w_k A)x\right]\\&=\frac{1}{m}\sum_{j=1}^m z_j^*[(I-w_j A)z_j],\quad z_j\equiv\prod_{\substack{k=1\\k\neq j}}^m(I-w_k A)x\\&=\frac{1}{m}\sum_{\substack{j=1\\z_j\neq 0}}^m\|z_j\|_2^2\left[1-w_j\left(\frac{z_j}{\|z_j\|_2}\right)^*A\left(\frac{z_j}{\|z_j\|_2}\right)\right].\end{aligned}$$

333

(e) 现在在(d)的恒等式中用$e^{i\theta}A$代替A，便得到对任意实数θ有

$$1-e^{im\theta}x^*A^m x=\frac{1}{m}\sum_{\substack{j=1\\z_j\neq 0}}^m\|z_j\|_2^2\left[1-e^{i\theta}w_j\left(\frac{z_j}{\|z_j\|_2}\right)^*A\left(\frac{z_j}{\|z_j\|_2}\right)\right].$$

然后假定$r(A)\leqslant 1$，证明这个恒等式右边的实部对任意$\theta\in\mathbf{R}$是非负的，由此推出左边的实部对所有$\theta\in\mathbf{R}$也一定是非负的．由于θ是任意的，证明这蕴涵$|x^*A^m x|\leqslant 1$，因而$r(A^m)\leqslant 1$.

28. 即使数值半径满足幂不等式$r(A^m)\leqslant r(A)^m$，但它对不等式$r(A^{k+m})\leqslant r(A^k)r(A^m)$不总成立．试通过考察$A=J_4(0)$（$4\times 4$ Jordan 块矩阵），$k=1$和$m=2$来验证这一事实．**提示**：

利用算术-几何平均值不等式证明，$r(A^2)=r(A^3)=\frac{1}{2}$，然后用 Cauchy-Schwarz 不等式证明 $r(A)<1$.

29. 类似于极小矩阵范数概念(5.6.29)，试问在 M_n 上存在关于"极小向量范数"的合理概念吗?

进一步阅读　关于涉及数值半径的不等式的进一步讨论可参看 M. Goldberg and E. Tadmor, "On the Numerical Radius and Its Applications," *Lin. Alg. Appl.* 42(1982), 263-284. 习题 27 中关于数值半径的幂不等式的证明取自 C. Pearcy, "An Elementary Proof of the Power Inequality for the Numerical Radius," *Michigan Math. J.* 13(1966), 289-291. 本节某些内容是本书作者的研究成果，可参看 C. R. Johnson, "Multiplicativity and Compatibility of Generalized Matrix Norms," *Linear Alg. Appl.* 16(1977), 25-37, "Locally Compatible Generalized Matrix Norms," *Numer. Math.* 27(1977), 391-394, "Power Inequalities and Spectral Dominance of Generalized Matrix Norms," *Linear Alg. Appl.* 28(1979), 117-130, 其中还有其他的结果.

[334]

5.8　矩阵的逆和线性方程组的解的误差

作为矩阵范数和向量范数的应用，我们来估计在求矩阵的逆和求线性方程组的解的过程中所引起的误差.

如果给定了非奇异矩阵 $A\in M_n$，可以设想能够准确地计算出逆矩阵 A^{-1}，但是，如果计算是在数字计算机上用有限长的计算机字符来完成，则不可避免地要出现舍入误差和舍位误差. 另外，即使能十分准确地完成所有计算，但可能有这种情形，矩阵 A 的有些元素是某些实验的结果或某些必然会出现误差的计算结果，因此，就不可能十分精确地知道这些结果. 那么，计算中的误差和数据中的误差如何影响要计算的逆矩阵呢?

对于许多普通的算法，可以证明，在计算中的舍入误差可以用同一种方式模拟成数据中的误差. 也就是说，如果假定 $A\in M_n$ 是给定的非奇异矩阵，并且想计算 A^{-1}，而实际计算的可能是$(A+E)^{-1}$，其中 $E\in M_n$ 是足够小的矩阵，且使 $A+E$ 为可逆矩阵. 于是，误差是 $A^{-1}-(A+E)^{-1}=A^{-1}-(I+A^{-1}E)^{-1}A^{-1}$. 如果$\rho(A^{-1}E)<1$，则 $A+E$ 就是可逆矩阵，且我们可以把$(I+A^{-1}E)^{-1}$写成 $A^{-1}E$ 的幂级数，这就给出

$$A^{-1}-(A+E)^{-1}=A^{-1}-\sum_{k=0}^{\infty}(-1)^k(A^{-1}E)^kA^{-1}=\sum_{k=1}^{\infty}(-1)^{k+1}(A^{-1}E)^kA^{-1}.$$

因此，关于这个误差的精确公式：

$$A^{-1}-(A+E)^{-1}=\sum_{k=1}^{\infty}(-1)^{k+1}(A^{-1}E)^kA^{-1},\text{如果 }\rho(A^{-1}E)<1. \tag{5.8.1}$$

现在假定$\|\cdot\|$是给定的矩阵范数，且假定$\|A^{-1}E\|<1$，因而特别有 $\rho(A^{-1}E)<1$ 和(5.8.1)成立. 于是

$$\|A^{-1}-(A+E)^{-1}\|=\left\|\sum_{k=1}^{\infty}(-1)^{k+1}(A^{-1}E)^kA^{-1}\right\|$$

$$\leqslant \sum_{k=1}^{\infty} \|A^{-1}E\|^k \|A^{-1}\| = \frac{\|A^{-1}E\|}{1-\|A^{-1}E\|}\|A^{-1}\|,$$

335

由此得出，求逆过程中所引起的相对误差的一个上界是

$$\frac{\|A^{-1}-(A+E)^{-1}\|}{\|A^{-1}\|} \leqslant \frac{\|A^{-1}E\|}{1-\|A^{-1}E\|}, \text{如果} \|A^{-1}E\| < 1. \tag{5.8.2}$$

另外，如果让 E 足够"小"，以至 $\|E\|<1/\|A^{-1}\|$，则 $\rho(A^{-1}E)\leqslant\|A^{-1}E\|\leqslant\|A^{-1}\|\|E\|<1$，且有估计式

$$\frac{\|A^{-1}-(A+E)^{-1}\|}{\|A^{-1}\|} \leqslant \frac{\|A^{-1}\|\|E\|}{1-\|A^{-1}\|\|E\|} = \frac{\|A^{-1}\|\|A\|(\|E\|/\|A\|)}{1-\|A^{-1}\|\|A\|(\|E\|/\|A\|)}.$$

数

$$\kappa(A) \equiv \begin{cases} \|A^{-1}\|\|A\| & \text{如果 } A \text{ 是非奇异矩阵；} \\ \infty & \text{如果 } A \text{ 是奇异矩阵} \end{cases} \tag{5.8.3}$$

称为矩阵逆关于矩阵范数$\|\cdot\|$的条件数. 注意 $\kappa(A)=\|A^{-1}\|\|A\|\geqslant\|A^{-1}A\|=\|I\|\geqslant1$ 对任意矩阵范数都成立.

利用这个记号，有估计式

$$\frac{\|A^{-1}-(A+E)^{-1}\|}{\|A^{-1}\|} \leqslant \frac{\kappa(A)}{1-\kappa(A)(\|E\|/\|A\|)}\frac{\|E\|}{\|A\|}, \text{如果} \|E\|\|A^{-1}\| < 1, \tag{5.8.4}$$

它用数据的相对误差表示矩阵逆的相对误差的界. 对小的$\|E\|$，右边为数量级 $\kappa(A)\|E\|/\|A\|$，所以有充足的理由相信，只要 $\kappa(A)$ 不很大，矩阵逆的相对误差与数据的相对误差就为同一个数量级，为了求逆的实际需要，如果 $\kappa(A)$ 较大，称 A(关于矩阵范数$\|\cdot\|$)为病态的或坏条件的；如果 $\kappa(A)$ 较小(接近于 1)，就称 A(关于矩阵范数$\|\cdot\|$)为良态的；如果 $\kappa(A)=1$，就称 A(关于范数$\|\cdot\|$)为优态的.

在采用的范数是谱范数的普通情形，条件数有一个有趣的几何特征. 设 $\theta(A)$ 表示当 x 和 y 取遍所有正交单位向量偶时向量 Ax 与 Ay 间的最小角. 利用谱范数，$\kappa(A)=\cot[\theta(A)/2]$. 因此，若 A 是酉矩阵，则 $\theta(A)=\pi/2$ 且 $\cot(\pi/4)=1=\kappa(A)$. 若 A"近乎奇异矩阵"，则有某个正交单位向量偶 x，y，使得 Ax"几乎平行"于 Ay，$\theta(A)$ 就会很小，而 $\kappa(A)=\cot[\theta(A)/2]$ 就会很大. 欲知详情，请参看例(7.4.26).

练习 证明，如果 $A\in M_n$ 可逆，则 $\kappa(A)=\kappa(A^{-1})$.

336

练习 证明，如果 U，$V\in M_n$ 是酉矩阵，且采用谱范数(或者任何其他酉不变范数)，则 $\kappa(A)=\kappa(UA)=\kappa(AV)=\kappa(UAV)$. 因此，一个给定矩阵的酉变换不会使它比原有的性态更加病态. 这个论断为许多稳定数值线性代数算法奠定了基础.

练习 证明总有 $\kappa(AB)\leqslant\kappa(A)\kappa(B)$. $\kappa(\cdot)$ 是 M_n 上的矩阵范数或向量范数吗?

利用上述这些条件同样可以对线性方程组解的精确性给出先验的界. 假定希望解方程组

$$Ax = b,\ A \in M_n,\ b \in \mathbf{C}^n \tag{5.8.5}$$

但是，因为计算上的误差或数据不精确，实际上是解方程组

$$(A+E)\hat{x}=b,\ A,E\in M_n,\ b\in \mathbf{C}^n. \tag{5.8.6}$$

关于误差 $x-\hat{x}$，可以说些什么呢？

如果 E 足够“小”，使得 $\rho(A^{-1}E)<1$，则根据(5.8.1)有

$$\begin{aligned} x-\hat{x} &= A^{-1}b-(A+E)^{-1}b=[A^{-1}-(A+E)^{-1}]b \\ &= \sum_{k=1}^{\infty}(-1)^{k+1}(A^{-1}E)^k A^{-1}b=\sum_{k=1}^{\infty}(-1)^{k+1}(A^{-1}E)^k x. \end{aligned}$$

如果$\|\cdot\|$是矩阵范数，且$\|A^{-1}E\|<1$，又如果$\|\cdot\|$是相容向量范数，则关于误差的范数的上界是

$$\|x-\hat{x}\|\leqslant\sum_{k=1}^{\infty}\|A^{-1}E\|^k\|x\|=\frac{\|A^{-1}E\|}{1-\|A^{-1}E\|}\|x\|.$$

关于相对误差，这说明，如果$\|A^{-1}E\|<1$，且$\|\cdot\|$是与矩阵范数$\|\cdot\|$相容的矩阵范数，则

$$\frac{\|x-\hat{x}\|}{\|x\|}\leqslant\frac{\|A^{-1}E\|}{1-\|A^{-1}E\|}. \tag{5.8.7}$$

值得注意的是，这与关于矩阵逆的相对误差的上界(5.8.2)是相同的，且与线性方程组的右边 b 无关.

利用 A 的条件数，用来推导(5.8.4)的证明同样可以证明，如果$\|A^{-1}\|\ \|E\|<1$，且向量范数$\|\cdot\|$与矩阵范数$\|\cdot\|$相容，则(5.8.5)的解的相对误差有界

337

$$\frac{\|x-\hat{x}\|}{\|x\|}\leqslant\frac{\kappa(A)}{1-\kappa(A)(\|E\|/\|A\|)}\frac{\|E\|}{\|A\|}. \tag{5.8.8}$$

无论采用什么算法来解线性方程组(5.8.5)，它的解的相对误差与其系数矩阵的逆的相对误差有相同的界.

像标准线性方程组(5.8.5)的系数矩阵 A 的元素一样，它的右边 b 的元素也可能产生误差，影响方程组的解，于是想用

$$(A+E)\hat{x}=b+e \tag{5.8.9}$$

代替(5.8.6)，其中 $E\in M_n$ 和 $e\in\mathbf{C}^n$ 看成是“小”的数据误差.

采用相同的方法，在关于(5.8.8)的相同假设下，(如果 $b\neq 0$)求得(5.8.5)的解与(5.8.9)的解之间的相对误差为

$$\frac{\|x-\hat{x}\|}{\|x\|}\leqslant\frac{\kappa(A)}{1-\kappa(A)(\|E\|/\|A\|)}\frac{\|E\|}{\|A\|}+\frac{\kappa(A)}{1-\kappa(A)(\|E\|/\|A\|)}\frac{\|e\|}{\|b\|} \tag{5.8.10}$$

这样，相对误差界有两项，一项是关于系数矩阵 A 的相对误差界，一项是关于右边 b 的相对误差界. 在确定关于解误差的界对数据误差的界的灵敏度方面，条件数 $\kappa(A)$ 仍然起着决定性的作用.

至此，我们的所有估计式都有一个先验的误差界；它们不涉及所计算出的解或者由它导出的任何量. 但是，假定通过计算已经求出方程组(5.8.5)的某个“解”$\hat{x}$. 恰好有 $A\hat{x}=b$ 的情形往往是不可能的，不过，可以从剩余向量 $r\equiv b-A\hat{x}$ 得到$\hat{x}$与真解 x 如何接近的估计式. 因为 $A^{-1}r=A^{-1}[b-A\hat{x}]=A^{-1}b-\hat{x}=x-\hat{x}$，所以有简明的界 $\|x-\hat{x}\|\leqslant\|A^{-1}r\|$. 如果$\|\cdot\|$是与

向量范数 $\|\cdot\|$ 相容的矩阵范数，则我们有 $\|b\|=\|Ax\|\leqslant\|A\|\|x\|$，或者当 $b\neq 0$ 时，$1\leqslant \|A\|\|x\|/\|b\|$，因而

$$\|x-\hat{x}\|\leqslant\|A^{-1}\|\|r\|\leqslant\frac{\|A\|\|x\|}{\|b\|}\|A^{-1}\|\|r\|=\|A\|\|A^{-1}\|\frac{\|r\|}{\|b\|}\|x\|.$$

因此，如果 $b\neq 0$，计算出的 $\hat{x}$(适合 $A\hat{x}=b-r$)与真解 x(适合 $Ax=b$)间的相对误差有界

$$\frac{\|x-\hat{x}\|}{\|x\|}\leqslant\kappa(A)\frac{\|r\|}{\|b\|},\tag{5.8.11}$$

其中，假定用来计算条件数 $\kappa(A)$ 的矩阵范数是与向量范数 $\|\cdot\|$ 相容的．对于良态问题，解的相对误差不会比剩余向量的相对大小(更)坏；对于病态问题，即使计算出的解产生一个小的剩余向量，它仍然可能与真解相差很大． 338

关于误差的范数估计的最后一个注释，我们指出，本节所导出的各个上界仅仅是上界而已．尽管上界可能很大，但是实际误差仍然可能很小．这些界的一个共同特征是它们的保守性：它们给出的误差界对许多问题未免太笼统了．但是，如果一个有适当阶数的矩阵 A 具有适当大小的各元素，且有较大的条件数，则 A^{-1} 一定有某些大元素，于是，对下述理由多加考虑是有益的．

如果 $Ax=b$，且令 $C=[c_{ij}]=A^{-1}$，那么，关于元 b_j 微分恒等式 $x=Cb$ 便得恒等式

$$\frac{\partial x_i}{\partial b_j}=c_{ij},\quad i,j=1,2,\cdots,n.\tag{5.8.12}$$

另外，把 $C=A^{-1}$ 看成 A 的函数，则它的各元只是 A 的各元的有理函数，因而可微．恒等式 $CA=I$ 表明，对所有 i，$q=1$，…，n，有

$$\sum_{p=1}^{n}c_{ip}a_{pq}=\delta_{iq},$$

因而

$$\sum_{p=1}^{n}\left[\frac{\partial c_{ip}}{\partial a_{jk}}a_{pq}+\delta_{pq,jk}c_{ip}\right]=\left[\sum_{p=1}^{n}\frac{\partial c_{ip}}{\partial a_{jk}}a_{pq}\right]+\delta_{qk}c_{ij}=0,$$

或

$$\sum_{p=1}^{n}\frac{\partial c_{ip}}{\partial a_{jk}}a_{pk}=-\delta_{qk}c_{ij},\quad i,j,k=1,\cdots,n.$$

现在关于 a_{jk} 微分恒等式 $x=Cb$ 使得

$$\begin{aligned}\frac{\partial x_i}{\partial a_{jk}}&=\sum_{p=1}^{n}\frac{\partial c_{ip}}{\partial a_{jk}}b_p=\sum_{p=1}^{n}\sum_{q=1}^{n}\frac{\partial c_{ip}}{\partial a_{jk}}a_{pq}x_q\\&=\sum_{q=1}^{n}\left[\sum_{p=1}^{n}\frac{\partial c_{ip}}{\partial a_{jk}}a_{pq}\right]x_q=\sum_{q=1}^{n}[-\delta_{qk}c_{ij}]x_q=-c_{ij}x_k,\end{aligned}$$

它是恒等式

$$\frac{\partial x_i}{\partial a_{jk}}=-c_{ij}\sum_{p=1}^{n}c_{kp}b_p.\tag{5.8.13}$$

因此，(5.8.12)和(5.8.13)同时提醒我们，如果 $C=A^{-1}$ 有相对大的任意元素，则解 x 的某个元素对 b 和 A 的某些元素的扰动就可能有难以避免的较大灵敏度． 339

习题

1. 证明非奇异正规矩阵的逆关于谱范数的条件数是

$$\kappa(A)=\rho(A)\rho(A^{-1})=|\lambda_{\max}(A)/\lambda_{\min}(A)|.$$

2. 求矩阵

$$A=\begin{bmatrix}1 & -1\\ -1 & 1+\varepsilon\end{bmatrix},\varepsilon>0$$

的特征值和逆矩阵. 证明，当 $\varepsilon\to0$ 时，A 的最大特征值与最小特征值之比为数量级 ε^{-1}. 试用习题1证明，关于谱范数有 $\kappa(A)=O(\varepsilon^{-1})$. 证明，关于任意范数有 $\kappa(A)=O(\varepsilon^{-1})$，因而对于求逆来说，$A$ 是病态的. 试用 A^{-1} 的显示形式证明，对任何范数有 $\kappa(A)=O(\varepsilon^{-1})$.

3. 求矩阵

$$B=\begin{bmatrix}1 & -1\\ 1 & 1+\varepsilon\end{bmatrix},\varepsilon>0$$

的特征值和逆矩阵. 证明，当 $\varepsilon\to0$ 时，B 的最大特征值与最小特征值之比为数量级1. 但是证明，关于任何矩阵范数有 $\kappa(B)=O(\varepsilon^{-1})$，因而当 $\varepsilon\to0$ 时，对于求逆来说，B 是病态的. 由此可知，最大模特征值与最小模特征值之比不一定是非正规矩阵的条件数.

4. 矩阵逆的条件数 $\kappa(A)$ 依赖于所采用的矩阵范数，但是可以证明，所有条件数在下述意义下是等价的：如果 $\kappa_\alpha(A)\equiv\|A^{-1}\|_\alpha\|A\|_\alpha$，且 $\kappa_\beta\equiv\|A^{-1}\|_\beta\|A\|_\beta$，则存在有正常数 C_m 和 C_M 使得

$$C_m\kappa_\alpha(A)\leqslant\kappa_\beta(A)\leqslant C_M\kappa_\alpha(A)$$

对所有 $A\in M_n$ 成立.

5. 证明，关于谱范数，每个酉矩阵 U 对于求逆是优态的[$\kappa(U)=1$]. 但是，如果采用 l_2 范数，每个酉矩阵 $U\in M_n$ 的条件数 $\kappa(U)$ 是 n.

6. 证明，对任意非奇异矩阵 $A\in M_n$ 和任意矩阵范数有 $\kappa(A)\geqslant|\lambda(A)|_{\max}/|\lambda(A)|_{\min}$（max 和 min 在这种情形是指特征值的极大模和极小模）. 因此，如果这个特征值的比是大的，则该矩阵对于求逆一定是病态的，而不管它是否为正规矩阵. 但是习题3说明，如果矩阵不是正规阵，即使这个比不大，它仍可能是病态的.

340

7. 试对界(5.8.8)的推广(5.8.10)给出详细证明，当 $e=0$ 时，界(5.8.10)就简化成(5.8.8).

8. 设 x 是 $\mathbf{C}^n$ 中的单位向量，且设 $\lambda>0$，证明 $A\equiv I+\lambda xx^*$ 是 Hermite 矩阵，且有 $(n-1)$ 重特征值1和特征值 $1+\lambda$. 证明，$\kappa(A)=1+\lambda$（关于谱范数），于是，这给出了产生可逆矩阵的简单方法，且使该矩阵具有有界元素和任意大的条件数. 为什么?

9. 设 B 是习题3中的矩阵，考虑具有精确解 $x=[1,0]^T$ 的线性方程组 $Bx=[1,1]^T$ 及其近似解 $\hat{x}=[1+\varepsilon^{-1/2},\varepsilon^{-1/2}]^T$. 证明，当 $\varepsilon\to0$ 时；剩余向量的相对误差为 $\|r\|/\|b\|=O(\varepsilon^{1/2})\to0$；但是，当 $\varepsilon\to0$ 时，解的相对误差为 $\|x-\hat{x}\|/\|x\|=O(\varepsilon^{-1/2})\to\infty$. 这样，有大误差的近似解可以产生小的剩余向量. 试用界(5.8.11)来说明.

10. 如果 $\det A$ 是小的(或大的)，$\kappa(A)$ 就一定大吗? **提示**：考虑 $A=\lambda I\in M_n$.

11. 一般，(5.8.4)的结果比(5.8.2)的要弱．这是因为假设条件$\|A^{-1}\|\|E\|<1$比$\|A^{-1}E\|<1$有更强的限制．不过，即使较强的假设条件被满足，(5.8.2)仍然可以给出比(5.8.4)较好的上界．试用$E=\varepsilon A$，$0<\varepsilon<1$，作出说明.

12. 如果方程(5.8.5)和(5.8.6)用

$$AX=B \quad 和 \quad (A+E)\hat{X}=B$$

来代替，其中，A，$E\in M_n$，而X，$B\in M_{n,k}$，求关于界(5.8.7)和(5.8.8)的类似界．考虑$k=n$和$B=I$的特殊情形；这有助于“说明”为什么(5.8.2)中与(5.8.7)中的上界是相同的吗?

13. 我们已经导出的矩阵逆的误差的各个界都依赖于(5.8.1)，它要求$\rho(A^{-1}E)<1$．证明，如果A和$A+E$都是可逆矩阵，则不论$A^{-1}E$的阶数如何，

$$\|A^{-1}-(A+E)^{-1}\|\leqslant\|A^{-1}\|\|(A+E)^{-1}\|\|E\|$$

对任何矩阵范数$\|\cdot\|$都成立．**提示：**证明$A^{-1}-(A+E)^{-1}=(A+E)^{-1}EA^{-1}$.

14. 或许被引用得最多的病态矩阵的例子是用$h_{ij}=1/(i+j-1)$定义的Hilbert矩阵$H_n=[h_{ij}]\in M_n$．证明H_n关于谱范数的条件数是用$\lambda_{\max}/\lambda_{\min}$给出的．其事实是$H_n$的条件数是渐近地等于$e^{cn}$，其中常数$c$大约是3.5，另一个事实是当$n\to\infty$时$\rho(H_n)=\pi+O[1/(\log n)]$．我们有$\kappa(H_3)\sim5\times10^2$，$\kappa(H_6)\sim1.5\times10^7$和$\kappa(H_8)\sim1.5\times10^{10}$．说明，即使$H_n$的各元都一致有界且$\rho(H_n)$不是很大，为什么$H_n$还是坏条件的. [341]

15. 如果采用谱范数，证明$\kappa(A^*A)=\kappa(AA^*)=[\kappa(A)]^2$．试说明，为什么采用数值方法解$A^*Ax=y$的问题可能不会比解$Ax=z$的问题真正来得容易.

16. 设$A\in M_n$是非奇异矩阵．试用(5.6)节37题中的不等式证明$\kappa(A)\geqslant\|A\|/\|A-B\|$对任一奇异矩阵$B\in M_n$成立．这里$\|\cdot\|$是任一矩阵范数，而$\kappa(\cdot)$是相应的条件数．这个下界可以用来证明一个给定矩阵$A$是病态的.

17. 设$A=[a_{ij}]\in M_n$是所有$a_{ii}\neq0$的上三角矩阵．试用习题16证明关于极大行和范数的条件数有下界

$$\kappa(A)\geqslant\frac{\|A\|}{\min\limits_{1\leqslant i\leqslant n}|a_{ii}|}.$$

进一步阅读 在解线性方程组时求误差的先验界的问题已是数值线性代数的一个中心问题；见[Ste]. [342]

第 6 章　特征值的估计和扰动

6.0　导引

对角矩阵的特征值是很容易确定的，并且矩阵的特征值是其诸元素的连续函数，于是，我们自然要问，如果矩阵的非对角元相对于主对角元是“小”的，那么，关于它的特征值，我们是否能够说出任何有用的结果来．这样的矩阵的确在实际中出现；由数值离散化椭圆型偏微分方程的边值问题得来的大线性方程组就可能具有这种形式．

在某些涉及一个振动系统的长期稳定性的微分方程问题中，人们感兴趣的往往是要证明一个矩阵的所有特征值都位于左半平面内，即 $\mathrm{Re}(\lambda_i)<0$．而在统计学或数值分析中，人们有时需要证明 Hermite 矩阵是正定的，即所有 $\lambda_i>0$．

有时我们想在一个容易刻化的有界集上估计一个矩阵的特征值．我们知道，矩阵 A 的所有特征值都位于复平面上中心在原点，半径为$\|A\|$的圆盘中，其中$\|\cdot\|$是任意矩阵范数．然而，能否用更准确的估计区域做出比这更好的估计，使得该区域肯定包含或不包含这些特征值呢？我们将会看到，这是能够做到的．

最后，假定确实知道 A 的各特征值，又假定 A 必然会产生扰动 $A\to A+E$．这时，特征值如何变化呢？因为 A 的特征值是 A 的各元的连续函数，所以有理由相信，如果扰动矩阵 E 有一个足够小的变化，则特征值不会有太大的变化．但是需要有准确的界限，以便知道在每种情形下怎样小才是所谓的“小”．这里的基本问题与(5.8)节中相同，在那里，我们讨论了线性方程组的解对数据扰动的灵敏度．

6.1　Geršgorin 圆盘

如果 $A\in M_n$，总可以把 A 写成 $A=D+B$，其中，$D=\mathrm{diag}(a_{11},\cdots,a_{nn})$正好是 A 的主对角部分，而 B 有零主对角线．如果对任意 $\varepsilon\in\mathbf{C}$，令 $A_\varepsilon\equiv D+\varepsilon B$，则 $A_0=D$，而 $A_1=A$．$A_0=D$ 的特征值是容易确定的：它们恰好是复平面中的点 a_{11}，…，a_{nn}．我们有理由推测，如果 ε 足够小，则 A_ε 的特征值将位于点 a_{11}，…，a_{nn}的某些小的邻域中．下述定理(常称为 Geršgorin 圆盘定理)使这个论断更为明确：确实存在某些以点 a_{ii} 为中心的容易计算的圆盘，它们肯定包含诸特征值．

6.1.1 定理(Geršgorin)　设 $A=[a_{ij}]\in M_n$，又设

$$R_i'(A)\equiv\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|,\quad 1\leqslant i\leqslant n$$

表示 A 的各去心绝对行和．则 A 的所有特征值都位于 n 个圆盘的并

$$\bigcup_{i=1}^{n}\{z\in\mathbf{C}:|z-a_{ii}|\leqslant R_i'(A)\}\equiv G(A)\tag{6.1.2}$$

中．此外，如果这 n 个圆盘中有 k 个之并形成一个连通区域，且它与所有余下的 $n-k$ 个圆盘

都不相交，则在这个区域中恰好有 A 的 k 个特征值.

证明：设 λ 是 A 的特征值，且假定 $Ax=\lambda x$，$x=[x_i]\neq 0$. x 有一个元素具有最大的绝对值，例如，对所有 $i=1,2,\cdots,n$，$|x_p|\geqslant|x_i|$，且 $x_p\neq 0$. 于是 $Ax=\lambda x$ 的假定说明

$$\lambda x_p=[\lambda x]_p=[Ax]_p=\sum_{j=1}^{n}a_{pj}x_j,$$

它等价于

344

$$x_p(\lambda-a_{pp})=\sum_{\substack{j=1\\ j\neq p}}^{n}a_{pj}x_j$$

另一方面，三角不等式使我们得出，

$$|x_p||\lambda-a_{pp}|=\left|\sum_{\substack{j=1\\ j\neq p}}^{n}a_{pj}x_j\right|\leqslant\sum_{\substack{j=1\\ j\neq p}}^{n}|a_{pj}x_j|=\sum_{\substack{j=1\\ j\neq p}}^{n}|a_{pj}||x_j|$$

$$\leqslant|x_p|\sum_{\substack{j=1\\ j\neq p}}^{n}|a_{pj}|=|x_p|R_p'.$$

因而，对某个 p，$|\lambda-a_{pp}|\leqslant R_p'$：即 λ 位于中心为 a_{pp}，半径为 R_p' 的闭圆盘中. 因为不知道哪个 p 适合每个特征值 λ(除非知道相应的特征向量，要是这样，我们就准确地知道了 λ，况且可能对确定 λ 不感兴趣)，所以只能得出，λ 位于所有这些圆盘的并中，它就是区域(6.1.2).

为了证明定理的第二个论断，记 $A=D+B$，其中 $D=\mathrm{diag}(a_{11},\cdots,a_{nn})$，且令 $A_\varepsilon\equiv D+\varepsilon B$. 注意，$R_i'(A_\varepsilon)=R_i'(\varepsilon B)=\varepsilon R_i'(A)$. 为方便起见，假定前 k 个圆盘

$$\bigcup_{i=1}^{k}\{z\in\mathbf{C}:|z-a_{ii}|\leqslant R_i'\}$$

形成连通区域 G_k，且它与余下的 $n-k$ 个圆盘组成的补区域 G_k' 不相交，即 $G_k'=G(A)\setminus G_k$. 值得指出的是，对所有 $\varepsilon\in[0,1]$，A_ε 的前 k 个圆盘的并

$$G_k(\varepsilon)\equiv\bigcup_{i=1}^{k}\{z\in\mathbf{C}:|z-a_{ii}|\leqslant R_i'(A_\varepsilon)=\varepsilon R_i'(A)\}$$

包含在连通集 $G_k\equiv G_k(1)$ 中，不过，对所有这样的 ε，$G_k(\varepsilon)$ 可能本身不是连通集. 另外，补区域 $G_k'(\varepsilon)\equiv G_n(\varepsilon)\setminus G_k(\varepsilon)$ 中没有一个与 G_k 相交. 对于每个 $i=1,\cdots,k$，考虑特征值 $\lambda_i(A_0)=a_{ii}$ 以及 $\lambda_i(A_\varepsilon)$，$\varepsilon>0$. 因为特征值是 A 诸元的连续函数(见附录 D)，又因为对所有 $\varepsilon\in[0,1]$ 都有 $\lambda_i(A_\varepsilon)\in G_k(\varepsilon)\subset G_k$，所以每个 $\lambda_i(A_0)$ 通过 G_k 中由 $\{\lambda_i(A_\varepsilon):0\leqslant\varepsilon\leqslant 1\}$ 给出的连续曲线与某个 $\lambda_i(A_1)=\lambda_i(A)$ 相连结. 对于每个 $\varepsilon\in[0,1]$，得知，在 $G_k(\varepsilon)$ 中至少包含 A_ε 的 k 个特征值. 但是，$G_k(\varepsilon)$ 中 A_ε 的特征值又不会多于 k 个，这是因为，A_0 的其余 $n-k$ 个特征值作为一些连续曲线的起点落在连通集 G_k 以外，而其终点肯定在补区域 G_k' 内；因为连续性和连通性(这是关于连续函数的中间值定理)，它们不可能跳越 G_k' 与 G_k 之间的空隙. □

(6.1.2)中的区域 $G(A)$ 常常称为 A 的(关于行的)Geršgorin 区域；$G(A)$ 中的各个圆盘称为 Geršgorin 圆盘，且这些圆盘的边界称为 Geršgorin 圆. 因为 A 和 A^T 有相同的特征值，为了得到用去心绝对列和

345

$$C_j'(A)=\sum_{\substack{i=1\\ i\neq j}}^{n}|a_{ij}|$$

描述的且包含 A 的各特征值的区域. 可以把 Geršgorin 圆盘定理应用于 A^T，得到关于列的 Geršgorin 圆盘定理.

6.1.3 推论 如果 $A=[a_{ij}]\in M_n$，则 A 的所有特征值都位于 n 个圆盘的并

$$\bigcup_{j=1}^{n}\{z\in \mathbf{C}: |z-a_{jj}|\leqslant C_j'\}=G(A^T) \tag{6.1.4}$$

中. 此外，如果这些圆盘中的 k 个之并形成一个连通区域，且与所有其余 $n-k$ 个圆盘不相交，则在这个区域内恰好有 A 的 k 个特征值.

练习 证明，A 的所有特征值都位于区域(6.1.2)和(6.1.4)的交中，也就是说，它们位于 $G(A)\cap G(A^T)$ 中. 试用具有 $a_{ij}=i/j$ 的 3×3 矩阵$[a_{ij}]$来说明.

因为 A 的所有特征值都位于(6.1.2)和(6.1.4)这两个区域中，所以 A 的最大模特征值也在其中. 在 $G(A)$的第 i 个圆盘中，离原点最远的点有模

$$|a_{ii}|+R_i'=\sum_{j=1}^{n}|a_{ij}|,$$

因而这些值的最大者一定是 A 的最大模特征值的一个上界. 当然，类似的论述也可对绝对列和作出.

6.1.5 推论 如果 $A=[a_{ij}]\in M_n$，则

$$\rho(A)\leqslant \min\left\{\max_i\sum_{j=1}^{n}|a_{ij}|,\ \max_j\sum_{i=1}^{n}|a_{ij}|\right\}.$$

这个结论是我们意料之中的，因为它是说，$\rho(A)\leqslant \|A\|_\infty$ 和$\|A^T\|_\infty$(极大绝对行和范数和极大绝对列和范数)，且这个不等式对任意矩阵范数成立. 但是，有趣的是，这个结论有一个本质的几何推论.

因为只要 S 是可逆矩阵，$S^{-1}AS$ 与 A 就有相同的特征值，所以可以把 Geršgorin 定理应用于 $S^{-1}AS$；也许对 S 的某个选择，所得到的界可能更准确. 一个特别方便的选择是 $S=D=\mathrm{diag}(p_1, p_2, \cdots, p_n)$，其中所有 $p_i>0$. 容易算出，$D^{-1}AD=[p_ja_{ij}/p_i]$. 把 Geršgorin 定理应用于 $D^{-1}AD$ 和它的转置便得到下面的推论. 346

6.1.6 推论 设 $A=[a_{ij}]\in M_n$，且设 $p_1, p_2, \cdots, p_n$ 是正实数. 则 A 的所有特征值位于区域

$$\bigcup_{i=1}^{n}\left\{z\in \mathbf{C}: |z-a_{ii}|\leqslant \frac{1}{p_i}\sum_{\substack{j=1\\ j\neq i}}^{n}p_j|a_{ij}|\right\}=G(D^{-1}AD)$$

中，同时也位于区域

$$\bigcup_{j=1}^{n}\left\{z\in \mathbf{C}: |z-a_{jj}|\leqslant p_j\sum_{\substack{i=1\\ i\neq j}}^{n}\frac{1}{p_i}|a_{ij}|\right\}=G[(D^{-1}AD)^T]$$

中.

矩阵 $A=\begin{bmatrix}1 & 1\\ 0 & 2\end{bmatrix}$有特征值 1 和 2. 直接应用 Geršgorin 定理只给出特征值的一个相当粗略的估计(图 6.1.7a)，而后一个论断中的备用参数却为得到特征值的一个令人满意的估计提供了很大的灵活性(图 6.1.7b).

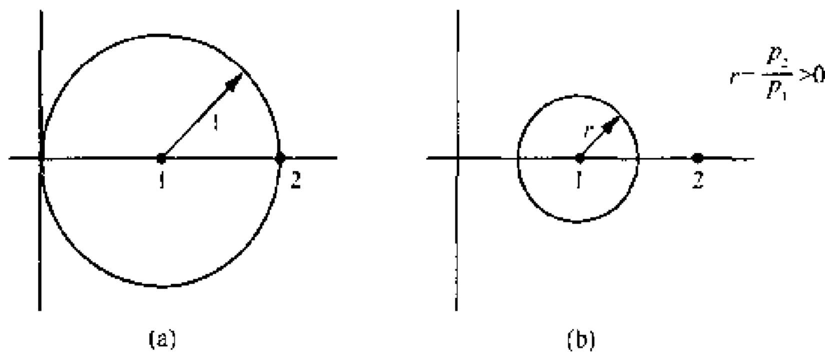


图 6.1.7

练习 考虑矩阵

$$A=\begin{bmatrix} 7 & -16 & 8 \\ -16 & 7 & -8 \\ 8 & -8 & -5 \end{bmatrix}.$$

[347]

试用 Geršgorin 定理估计一下 A 的特征值以及 A 的谱半径. 然后考虑 $D^{-1}AD$, 其中, $D=\text{diag}(p_1, p_2, p_3)$. 并且在你的特征值估计中, 看一看能否作出什么改进, 最后, 计算出实际的特征值, 试对你所作估计的好坏给一个评价.

练习 证明, A 的每个特征值位于集合 $\bigcap_D G(D^{-1}AD)$ 中, 其中交取遍只具有正主对角元的所有对角矩阵.

我们也可以利用引进自由参数的思想得到关于谱半径估计(6.1.5)的更一般的形式.

6.1.8 推论 设 $A=[a_{ij}]\in M_n$. 则

$$\rho(A)\leqslant \min_{p_1,\cdots,p_n>0}\ \max_{1\leqslant i\leqslant n}\frac{1}{p_i}\sum_{j=1}^{n}p_j\,|a_{ij}|,$$

且

$$\rho(A)\leqslant \min_{p_1,\cdots,p_n>0}\ \max_{1\leqslant j\leqslant n}\ p_j\sum_{i=1}^{n}\frac{1}{p_i}\,|a_{ij}|.$$

练习 证明推论(6.1.8).

练习 设 $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, 其中 a, b, c 和 d 恒为实正数.

(a) 用直接计算求出一个具体的对角矩阵 $\widetilde{D}$, 使得 $\|\widetilde{D}^{-1}A\widetilde{D}\|_\infty=\min_D\|D^{-1}AD\|_\infty$, 其中极小取遍只具有正主对角元的所有对角矩阵.

(b) 计算 $\|\widetilde{D}^{-1}A\widetilde{D}\|_\infty=r$.

(c) 直接计算 $\rho(A)$.

(d) 指出 $r=\rho(A)$.

以后将证明, 如果 A 是任意一个 $n\times n$ 正矩阵(或更一般地为不可约矩阵和非负矩阵), 则 $D^{-1}AD$ 的极大行和在所有 D 上取极小一定等于谱半径. 这对一般情形不成立.

练习 考虑 $A=\begin{bmatrix}1 & 1\\ -1.5 & 2\end{bmatrix}$，证明，对具有正 p_1 和 p_2 的所有 $D=\mathrm{diag}(p_1, p_2)$，$\rho(A)<\min\|D^{-1}AD\|$.

如果有关于一个矩阵的某些附加信息，它要求特征值位于(或不位于)某些集合中，那么，可以利用这个信息以及 Geršgorin 定理给出特征值的更为准确的估计．例如，如果 A 是 Hermite 矩阵，则 A 的特征值一定都是实的，因而它们一定位于集 $\mathbf{R}\cap G(A)$ 中，它是实闭区间的有限并.

348

练习 关于斜 Hermite 矩阵的特征值的估计你能说些什么？关于酉矩阵呢？关于实正交矩阵呢？

因为一个矩阵是可逆的当且仅当 0 不是其特征值，所以，值得研究的是导出从包含特征值的已知区域中去掉原点的条件.

6.1.9 定义 设 $A=[a_{ij}]\in M_n$. 矩阵 A 称为对角占优的，是指

$$|a_{ii}|\geqslant\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|=R_i'$$

对所有 $i=1,\cdots,n$ 成立．称 A 为严格对角占优的，是指

$$|a_{ii}|>\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|=R_i'$$

对所有 $i=1,\cdots,n$ 成立.

从位置几何学可知，如果 A 是严格对角占优的，0 不可能位于任何闭 Geršgorin 圆盘中．另外，如果所有主对角元 a_{ii} 都是正实数，则每个圆盘实际上位于开右半平面内；如果 A 还是 Hermite 矩阵，则特征值一定都是正数．把这些论断总结为下述定理，其中(a)部分称为 Levy-Desplangues 定理[见推论(5.6.17)].

6.1.10 定理 设 $A=[a_{ij}]\in M_n$ 是严格对角占优的．那么

(a) A 是可逆矩阵.

(b) 如果 A 的所有主对角元都是正数，则 A 的所有特征值都有正实部.

(c) 如果 A 是 Hermite 矩阵，且 A 的所有主对角元都是正数，则 A 的所有特征值都是正实数.

练习 考察 $\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$和$\begin{bmatrix}1 & 1\\ 1-\varepsilon & 1\end{bmatrix}$，说明仅有对角占优性质还不足以确保可逆性，并且说明严格对角占优性对可逆性不是必要条件.

349

利用推论(6.1.6)的备用参数，作为可逆性的充分条件的严格对角占优性假设可以稍微放宽一些.

6.1.11 定理 设 $A=[a_{ij}]\in M_n$ 的所有对角元非零，且 A 是对角占优的，又除了 $i=1,\cdots,n$ 的一个值以外，对所有其他的值有 $|a_{ii}|>R_i'$. 则 A 是可逆矩阵.

证明： 假设条件是说，对某个 k，$|a_{kk}|=R_k'$，且对所有 $i\neq k$，$|a_{ii}|>R_i'$. 在(6.1.6)

中，对所有 $i\neq k$ 设 $p_i=1$，且设 $p_k=1+\varepsilon$，$\varepsilon>0$. 于是，对任意 $\varepsilon>0$，

$$\frac{1}{p_k}\sum_{\substack{j=1\\j\neq k}}^{n}p_j\mid a_{kj}\mid=\frac{1}{1+\varepsilon}R'_k<\mid a_{kk}\mid,$$

且对所有 $i\neq k$，

$$\frac{1}{p_i}\sum_{\substack{j=1\\j\neq i}}^{n}p_j\mid a_{ij}\mid=R'_i+\varepsilon\mid a_{ik}\mid.$$

但是，因为 $R'_i<\mid a_{ii}\mid$ 对所有 $i\neq k$ 成立，所以选择充分小的 $\varepsilon>0$，使得 $R'_i+\varepsilon\mid a_{ik}\mid<\mid a_{ii}\mid$ 对所有 $i\neq k$ 成立. 因此，根据推论(6.1.6)，点 $z=0$ 在 $G(D^{-1}AD)$之外，因而 A 必定是可逆矩阵. □

Geršgorin 定理及其各种变形给出了 A 的特征值的各种包含区域，这些区域只依赖于 A 的诸主对角元以及各非主对角元的绝对值. 利用 $S^{-1}AS$ 与 A 有相同的特征值的事实，可以得出(6.1.6)，并且得出闭集

$$\bigcap_D G(D^{-1}AD),\quad D=\operatorname{diag}(p_1,\cdots,p_n),\quad 所有\ p_i>0 \tag{6.1.12}$$

包含 $A\in M_n$ 的所有特征值的结论. 我们知道，如果可以采用比对角相似更复杂的相似，就有可能得到包含特征值的更小的区域，但是，如果只局限于对角相似，并且只限于利用主对角元以及非对角元的绝对值，能否设法得到比(6.1.12)更好的估计呢?

回答是否定的，其理由如下：设 z 是集合(6.1.12)的边界上任一给定的点. 那么，R. Varga 已经证明，存在矩阵 $B=[b_{ij}]\in M_n$，使得对所有 $i=1,\cdots,n$ 有 $b_{ii}=a_{ii}$，而对所有 $i,j=1,\cdots,n$ 有 $\mid b_{ij}\mid=\mid a_{ij}\mid$，并且 z 是 B 的特征值.

习题

350

1. 考虑 $n\times n$ 线性方程组 $Ax=y$，其中 A 和 y 是给定的：

(ⅰ) 定义 $B\equiv I-A$，然后把原方程组改写成 $x=Bx+y$.

(ⅱ) 按你所需要的任一方式选择解的初始逼近 $x^{(0)}$.

(ⅲ) 对 $m=0,1,2,\cdots$，计算 $x^{(m+1)}=Bx^{(m)}+y$.

(ⅳ) 可以指望，当 $m\to\infty$时，$x^{(m)}\to x$(解).

(a)试用 $\varepsilon^{(m)}=x^{(m)}-x$ 表示第 m 次逼近与解的误差，并且证明 $\varepsilon^{(m)}=B^m(x^{(0)}-x)$. (b)证明，如果 $\rho(I-A)<1$，则不论初始逼近 $x^{(0)}$ 如何选择，在 $m\to\infty$时 $x^{(m)}\to x$ 的意义下这个算法总是可行的. (c)试用 Geršgorin 定理给出关于 A 的一个简单明确的条件以保证这个算法可行.

2. 证明，$\bigcap_S G(S^{-1}AS)=\sigma(A)$，只要这个交取遍所有非奇异矩阵 S.

3. 利用(6.1.5)证明，对任意 $A\in M_n$，

$$\mid\det A\mid\leqslant\prod_{i=1}^{n}\Big(\sum_{j=1}^{n}\mid a_{ij}\mid\Big),$$

且对列也有类似的不等式. **提示**：如果 A 的任意一行是零，那就无需证明. 如果 A 的所有行都非零，则设 B 是这样的矩阵，它的各行是 A 的对应行除以该行的绝对行和，于是，根据(6.1.5)，$\rho(B)\leqslant1$，因而 $\mid\det B\mid\leqslant1$. 注意，这说明，

$$|\det A| \leqslant \prod_{i=1}^{n} \|a_i\|_1,$$

其中向量 a_i 是 A 的相应行(或列). 对其他范数，存在这样的不等式吗？**提示**：见(7.8.2).

4. 在正文中，由 Geršgorin 定理(6.1.1)导出了定理(6.1.10a)-Levy-Desplanques 定理. 证明，由(6.1.10)的(a)部分可推出(6.1.1)的前一部分[区域(6.1.2)包含 A 的所有特征值的论断]. **提示**：把(6.1.10a)应用于矩阵 $\lambda I - A$.

5. 假定 $A \in M_n$ 是实矩阵，且它的 n 个 Geršgorin 圆盘都彼此分离. 证明 A 的所有特征值都是实数. 更一般地，如果复矩阵 $A \in M_n$ 只有实主对角元，且它的特征多项式只有实系数，又它的 n 个 Geršgorin 圆盘都彼此分离，证明 A 的所有特征值都是实数.

6. 证明，如果 $A=[a_{ij}] \in M_n$，且对于 i 的 k 个不同的值有 $|a_{ii}| > R_i'$，则 $k \leqslant \text{rank } A$.

7. 假定 $A \in M_n$ 是幂零的($A^2=A$)但是 $A \neq I$. 证明 A 不可能是严格对角占优矩阵[或不可约对角占优矩阵；见(6.2.25)和(6.2.27)]. [351]

8. 假定 $A \in M_n$ 是严格对角占优矩阵，即 $|a_{ii}| > R_i'$ 对所有 $i=1, \cdots, n$ 成立. 证明 $|a_{kk}| > C_k'$ 对 $k=1, \cdots, n$ 的至少一个值成立.

9. 假定 $A=[a_{ij}] \in M_n$ 是严格对角占优矩阵，且设 $D=\text{diag}(a_{11}, a_{22}, \cdots, a_{nn})$，证明，$D$ 是可逆矩阵，且 $\rho(I-D^{-1}A)<1$. **提示**：利用推论(6.1.5).

10. 设 $A=[a_{ij}] \in M_n$，且 $R_i = R_i' + |a_{ii}|$ 表示 A 的第 i 行所有元素的绝对值的和，证明

$$\text{rank } A \geqslant \sum_{i=1}^{n} \frac{|a_{ii}|}{R_i},$$

其中约定在这个和中 $0/0=0$. **提示**：用同一个非零纯量乘一行的所有元素不改变矩阵的秩，因而可以假定所有 $a_{ii} \geqslant 0$ 且所有 R_i 或者是零或者是 1. 在这种情形，A 的所有特征值都位于单位圆盘内且可以证明

$$\text{rank } A \geqslant \sum_{i=1}^{n} a_{ii}.$$

证明，$\sum a_{ii} = \text{tr } A = \sum \lambda_i \leqslant \sum |\lambda_i| \leqslant A$ 的非零特征值的个数 $\leqslant \text{rank } A$.

11. 如果 $A=[a_{ij}]=[a_1\ a_2 \cdots a_n] \in M_n$，证明

$$\text{rank } A \geqslant \sum_{i=1}^{n} \frac{|a_{ii}|^2}{\|a_i\|_2^2},$$

其中约定在这个和中 $0/0=0$. **提示**：如同习题 10，说明只需考虑下述情形就可以了：A 的所有列有 Euclid 单位长度，即所有 $\|a_i\|_2=1$；且在这种情形可以证明

$$\text{rank } A \geqslant \sum_{i=1}^{n} |a_{ii}|^2 = \sum_{i=1}^{n} |e_i^* a_i|^2,$$

其中$\{e_1, e_2, \cdots, e_n\}$是 $\mathbf{C}^n$ 的标准正交基. 如果 A 有秩 k，证明存在 k 个标准正交向量 $v_1, \cdots, v_k \in \mathbf{C}^n$ 使得 $\text{Span}\{v_1, \cdots, v_k\} = \text{Span}\{a_1, \cdots, a_n\}$. 于是

$$a_i = \sum_{j=1}^{k} (v_j^* a_i) v_j, \quad 因而 \quad e_i^* a_i = \sum_{j=1}^{k} (v_j^* a_i)(e_i^* v_j),$$

且 [352]

$$\sum_{i=1}^{n}|e_i^*a_i|^2\leqslant\sum_{i=1}^{n}\Big[\Big(\sum_{j=1}^{k}|v_j^*a_i|^2\Big)\Big(\sum_{j=1}^{k}|e_i^*v_j|^2\Big)\Big]=\sum_{j=1}^{k}\sum_{i=1}^{n}|e_i^*v_j|^2=\sum_{j=1}^{k}1$$
$$=k=\text{rank }A.$$

进一步阅读 用数值例子讨论 Geršgorin 定理可以在[Ste]中找到. 原始资料可参看 S. Geršgorin, "Über die Abgrenzung der Eigenwerte einer Matrix," *Izv. Akad. Nauk. S. S. S. R.* 7(1931), 749-754. Geršgorin 定理有一个推广, 它给出了关于广义特征值问题 $Ax=\lambda Bx$ 的谱的包含区域, 其中包括 B 是奇异矩阵的情形; 可参看 G. W. Stewart, "Gerschgorin Theory for the Generalized Eigenvalue Problem," *Math. Comput.* 29(1975), 600-606. 关于本节最后一段提到区域(6.1.12)是否有更好的性质, 可参看 R. Varga, "Minimal Gerschgorin Sets," *Pacific J. Math.* 15(1965), 719-729.

6.2 Geršgorin 圆盘——更细致的讨论

我们已经看到, 严格对角占优性对于可逆性是充分的, 但对角占优性则不是. 某些 2×2 的矩阵例子所需条件启发我们猜测, 对角占优性连同严格不等式

$$|a_{ii}|>\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|\quad\text{对 } i=1,\cdots,n \text{ 中至少一个值成立}\tag{6.2.1}$$

可能是可逆性的充分条件. 令人失望的是, 正如例子

$$\begin{bmatrix}4&2&1\\0&1&1\\0&1&1\end{bmatrix}\tag{6.2.2}$$

所说明的那样, 情况并非如此. 这里究竟会出现什么情形呢?

但是, 关于对角占优矩阵, 存在一些普通条件, 在这些条件下, (6.2.1)足以保证可逆性, 并且它们在图论中诱导出一些很有意义的想法. 一个重要论断是, 如果 A 是对角占优矩阵, 则 0 不可能是任何单个 Geršgorin 圆盘的内点.

353

练习 证明, 如果一个给定点 λ 不是 A 的任一 Geršgorin 圆盘的内点, 当且仅当

$$|\lambda-a_{ii}|\geqslant R_i'=\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|\tag{6.2.2a}$$

对所有 $i=1,\cdots,n$ 成立. 证明 $G(A)$ 边界上的每一点 λ 适合这些不等式. 考察 $\lambda=0$ 和 $A=\begin{bmatrix}1&1\\1&i\end{bmatrix}\oplus\begin{bmatrix}-1&1\\1&-i\end{bmatrix}$ 来证明, $G(A)$ 的一个内点也可能满足不等式(6.2.2a).

仔细分析定理(6.1.1)的证明, 当 A 的特征值满足不等式(6.2.2a)时会出现什么情形.

6.2.3 引理 设 $A=[a_{ij}]\in M_n$, λ 是 A 的位于 $G(A)$ 的边界上的特征值. 设 $Ax=\lambda x$, $x=[x_i]\neq 0$, 且假定 p 是使 $|x_p|=\max_{1\leqslant i\leqslant n}|x_i|=\|x\|_\infty\neq 0$ 的一个下标. 那么,

(a) 如果 k 是使 $|x_k|=|x_p|$ 的任一个下标, 则 $|\lambda-a_{kk}|=R_k'$; 即第 k 个 Geršgorin 圆经过 λ;

(b) 如果对某个 $k=1,\cdots,n$, $|x_k|=|x_p|$, 又如果对某个 $j\neq k$, $a_{kj}\neq 0$, 则也有

$|x_j|=|x_p|$.

证明：正如在 Geršgorin 定理的证明中所证明的，有

$$(\lambda-a_{ii})x_i=\sum_{\substack{j=1\\j\neq i}}^{n}a_{ij}x_j$$

对所有 $i=1,\cdots,n$ 成立，因而

$$|\lambda-a_{ii}||x_i|=\left|\sum_{\substack{j=1\\j\neq i}}^{n}a_{ij}x_j\right|\leqslant\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}x_j|=\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_j|$$

$$\leqslant\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_p|=R_i'|x_p|. \tag{6.2.4}$$

因此，如果 k 是使 $|x_k|=|x_p|$ 的任一下标，一定有 $|\lambda-a_{kk}|\leqslant R_k'$. 但是，假设条件是，对所有 $i=1,\cdots,n$ 有 $|\lambda-a_{ii}|\geqslant R_i'$. 因此，对 $i=k$，在(6.2.4)的不等式两边，实际上应取等式；即

$$|\lambda-a_{kk}||x_k|=\sum_{\substack{j=1\\j\neq k}}^{n}|a_{kj}||x_j|=\sum_{\substack{j=1\\j\neq k}}^{n}|a_{kj}||x_k|=R_k'|x_k|. \tag{†}$$

因为 $|x_k|=\|x\|_\infty>0$，所以论断(a)可以从恒等式

$$|\lambda-a_{kk}||x_k|=R_k'|x_k|$$

推出. 论断(b)可从(†)中的中间恒等式

$$\sum_{\substack{j=1\\j\neq k}}^{n}|a_{kj}|(|x_k|-|x_j|)=0$$

推出，因为在这个和中每一项必是非负的. □

354

这个引理看起来更偏重于技巧，但是它有下述有用的结果及其推论作为直接推论.

6.2.5 定理 设 $A\in M_n$，又设 λ 是 A 的特征值，且 λ 是 $G(A)$ 的边界点，或更一般地，λ 满足不等式(6.2.2a)，假定 A 的所有元都是非零元. 那么

(a) A 的每个 Geršgorin 圆经过 λ；

(b) 如果 $Ax=\lambda x$，$x=[x_i]\neq 0$，则对所有 $i,j=1,\cdots,n$ 有 $|x_i|=|x_j|$.

练习 从引理(6.2.3)推导定理(6.2.5).

6.2.6 推论 设 $A=[a_{ij}]\in M_n$，且假定 A 的所有元都是非零元. 如果 A 是对角占优的，且对 $i=1,\cdots,n$ 的至少一个值有 $|a_{ii}|>R_i'$，则 A 是可逆矩阵.

证明：假如 A 不是可逆矩阵，则 0 就是 A 的一个特征值. 因为 A 是对角占优矩阵，所以 0 不可能是 $G(A)$ 的内点，因而它一定是一个边界点. 定理(6.2.5)说明，每个 Geršgorin 圆必定经过 0，但是，如果 $|a_{ii}|>R_i'$，则第 i 个圆不可能经过 0. □

前一个结果既实用，也有意义，不过，如果利用引理(6.2.3)中更细致的结果，还可以得到更好的结论(不涉及 A 中有关零元的假设条件).

6.2.7 定义 我们称矩阵 $A=[a_{ij}]\in M_n$ 具有性质 SC，是指对适合 $1\leqslant p,q\leqslant n$ 的每对不同的整数 p,q，存在一系列互不相同的整数 $k_1=p,k_2,k_3,\cdots,k_{m-1},k_m=q$，$1\leqslant m\leqslant n$，使得所

有矩阵元素 $a_{k_1k_2}$，$a_{k_2k_3}$，…，$a_{k_{m-1}k_m}$ 都是非零元.

例如，矩阵(6.2.2)不具有性质 SC，因为数对 2，1 不容许有这样的非零元序列. 但是数对 1，2 的确有这样的序列.

355 利用这个概念和引理(6.2.3)，可以得到关于(6.2.5)的下述改进.

6.2.8 较好定理 设 $A=[a_{ij}]\in M_n$，又假定 λ 是 A 的特征值，且 λ 还是 $G(A)$ 的边界点，或更一般地，λ 满足不等式(6.2.2a). 如果 A 有性质 SC，那么，

(a) 每个 Geršgorin 圆经过 λ；

(b) 如果 $Ax=\lambda x$，且 $x=[x_i]\neq 0$，则对所有 $i,j=1,\cdots,n$ 有 $|x_i|=|x_j|$.

证明：设 $Ax=\lambda x$，且对所有 $i=1,\cdots,n$ 有 $|x_i|\leqslant|x_p|=\|x\|_\infty>0$. 于是由引理(6.2.3)，$|\lambda-a_{pp}|=R'_p$. 设 q 是任一其他下标，$1\leqslant q\leqslant n$，$q\neq p$. 因为 A 有性质 SC，所以存在一系列互不相同的下标 $k_1=p$，k_2，k_3，…，$k_m=q$，使得所有矩阵元素 $a_{k_1k_2}$，…，$a_{k_{m-1}k_m}$ 都非零. 因为 $a_{k_1k_2}=a_{pk_2}\neq 0$，根据(6.2.3)的论断(b)可知，$|x_p|=|x_{k_2}|$. 但另一方面，$a_{k_2k_3}\neq 0$，因而 $|x_{k_3}|=|x_{k_2}|=|x_p|$. 继续这样做下去，得知，对所有 $i=1,\cdots,m$ 有 $|x_{k_i}|=|x_p|$，因而[由(6.2.3)(a)]有，$|\lambda-a_{k_mk_m}|=|\lambda-a_{qq}|=R'_q$；即第 q 个 Geršgorin 圆经过 λ 且 $|x_q|=|x_p|$. 但因为 q 是任意下标，得知每个 Geršgorin 圆经过 λ 且所有 $|x_i|=|x_p|$，$i=1,\cdots,n$. □

如同(6.2.6)中那样，从这个结果可以推导出一个关于可逆性的有用充分条件.

6.2.9 较好推论 设 $A=[a_{ij}]\in M_n$，且假定 A 有性质 SC. 如果 A 是对角占优的，又如果对 $i=1,\cdots,n$ 的至少一个值有 $|a_{ii}|>R'_i$，则 A 是可逆矩阵.

练习 由(6.2.8)推导(6.2.9).

练习 证明矩阵(6.2.2)不具有性质 SC.

这个陌生的性质 SC 指的是什么？注意到，它只涉及 A 的非零非对角元的位置，不涉及主对角元，且非主对角元的具体值是无关紧要的. 由于这个缘故，我们定义与 A 有关的两个矩阵.

6.2.10 定义 如果 $A=[a_{ij}]\in M_{m,n}$，令 $|A|=[|a_{ij}|]$ 和 $M(A)=[\mu_{ij}]$，其中，如果 $a_{ij}\neq 0$，则 $\mu_{ij}=1$，如果 $a_{ij}=0$，则 $\mu_{ij}=0$. 矩阵 $M(A)$ 称为 A 的指标矩阵.

练习 证明，矩阵 $A\in M_n$ 有性质 SC，当且仅当 $|A|$ 或者 $M(A)$（因而两者都）有性质 SC.

356

在叙述性质 SC 过程中出现了 A 的非零元序列概念，它可以用与 A 相关联的一个图中的某些确定道路来形象地描述.

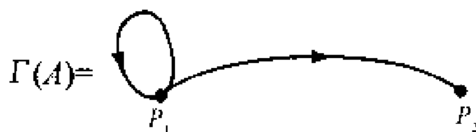
6.2.11 定义 $A\in M_n$ 的有向图，记作 $\Gamma(A)$，是关于 n 个结点 P_1，P_2，…，P_n 的有向图，且有以下性质：在 $\Gamma(A)$ 中存在一条从 P_i 到 P_j 的有向弧，当且仅当 $a_{ij}\neq 0$（$\mu_{ij}\neq 0$）.

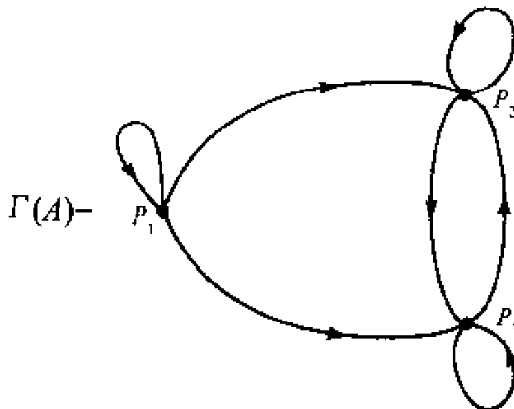
例

$$A=\begin{bmatrix}1&1\\1&1\end{bmatrix};$$

$\Gamma(A)=$ 

$A=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$; $\Gamma(A)=$ 

$A=\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}$; $\Gamma(A)=$ 

$A=\begin{bmatrix}4 & 2 & 1\\ 0 & 1 & 1\\ 0 & 1 & 1\end{bmatrix}$; $\Gamma(A)=$ 

6.2.12 定义 图 Γ 中的有向道路 γ 是 Γ 中的一个弧序列 $P_{i_1}P_{i_2}$，$P_{i_2}P_{i_3}$，$P_{i_3}P_{i_4}$，…. 有向道路 γ 中的有序结点表是 P_{i_1}，P_{i_2}，…. 如果在有向道路中，依次出现的弧的数目是有限的，就称这个数为该有向道路的长；否则，就称有向道路有无限长. 一个*回路*是起点和终点在同一个结点的有向道路；这个结点在该道路的有序结点表中恰好出现两次，且在该表中，没有其他结点出现两次以上；有些作者称这个回路为*简单有向回路*. 长为 1 的回路称为*圈*或*平凡回路*. 357

6.2.13 定义 有向图 Γ 是*强连通的*，是指在 Γ 中的每对不同结点 P_i，P_j 之间都存在一条以 P_i 为起点，P_j 为终点的有向道路，且有有限长.

6.2.14 定理 设 $A\in M_n$. 那么，A 有性质 SC，当且仅当有向图 $\Gamma(A)$ 是强连通的.

练习 证明定理 6.2.14.

练习 证明如果 Γ 具有性质：每对结点至少属于一个回路，则 Γ 是强连通的，但是，逆命题不成立. **提示**：

$$\begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}.$$

在有向图的两个结点之间，可能存在多条有向道路，但是具有不同长度的这样两条道路可能没有本质的差别；它们可能包含一条或多条重复出现的子道路. 显然，如果我们沿着一条有向道路前进时，曾两次遇到某个结点，那么去掉第一次遇到该结点到第二次遇到该结点间的所有中间弧(去掉的子图是一条回路或包含一条回路)，这条有向道路便可以缩短(且两个端点不受影响).

6.2.15 论断 设 Γ 是 n 个结点上的有向图. 如果在两个给定的结点之间存在 Γ 中的一条有向道路，则在这两个结点之间存在一条长不超过 $n-1$ 的有向道路.

我们应该怎样说明给定的矩阵 A 是否具有性质 SC 呢？这等价于验证 $\Gamma(A)$ 是否是强连通

的. 如果 n 不很大，或者如果 $M(A)$ 有一个特殊的结构，那么可以只检查 $\Gamma(A)$ 并划出所有可能的结点对之间的道路. 但是，这在一般情形下是不实用的，因此需有某个明确的计算方法.

6.2.16 定理 设 $A\in M_n$ 是给定的矩阵，又设 P_i 和 P_j 是 $\Gamma(A)$ 的给定结点. 那么，在 P_i 和 P_j 之间存在 $\Gamma(A)$ 中的一条长为 m 的有向道路，当且仅当 $(|A|^m)_{ij}\neq 0$，或等价地，如果 358 $[M(A)^m]_{ij}\neq 0$.

证明：我们用归纳法证明. 对 $m=1$，结论是明显的. 对 $m=2$，算出

$$[|A|^2]_{ij}=\sum_{k=1}^{n}[|A|]_{ik}[|A|]_{kj}=\sum_{k=1}^{n}|a_{ik}||a_{kj}|,$$

因而，$[|A|^2]_{ij}\neq 0$，当且仅当对 k 的至少一个值，a_{ik} 和 a_{kj} 都是非零的. 但是，这种情况成立，当且仅当在 $\Gamma(A)$ 中存在一条从 P_i 到 P_j 的道路. 一般地，假定对 $m=q$ 结论已经证明. 于是

$$[|A|^{q+1}]_{ij}=\sum_{k=1}^{n}[|A|^q]_{ik}[|A|]_{kj}=\sum_{k=1}^{n}[|A|^q]_{ik}|a_{kj}|\neq 0$$

当且仅当对 k 的至少一个值，$[|A|^q]_{ik}$ 和 $|a_{kj}|$ 都是非零的. 这等价于有一条从 P_i 到 P_k 的长为 q 的道路及一条从 P_k 到 P_j 的长为 1 的道路，而这种情况成立，当且仅当存在一条从 P_i 到 P_j 的长为 $q+1$ 的道路. 同样的证明对 $M(A)$ 也是可行的. □

6.2.17 定义 设 $A=[a_{ij}]\in M_n$. 我们说 $A\geqslant 0$（A 是非负矩阵），是指它的元素 a_{ij} 都是非负实数. 我们说 $A>0$（A 是正矩阵），是指它的元素 a_{ij} 都是正实数.

6.2.18 推论 设 $A\in M_n$. 那么，$|A|^m>0$，当且仅当从 $\Gamma(A)$ 中的每个结点 P_i 到每个结点 P_j，在 $\Gamma(A)$ 中都存在一条长恰好为 m 的有向道路. 相同的结论对 $M(A)^m$ 也成立.

6.2.19 推论 设 $A\in M_n$. 那么，A 有性质 SC，当且仅当 $(I+|A|)^{n-1}>0$，或等价地，如果 $[I+M(A)]^{n-1}>0$.

证明：$(I+|A|)^{n-1}=I+(n-1)|A|+\binom{n-1}{2}|A|^2+\cdots+\binom{n-1}{n-2}|A|^{n-1}>0$，当且仅当对诸结点的每对 (i,j)，$i\neq j$，在诸项 $|A|$，$|A|^2$，…，$|A|^{n-1}$ 中，至少有一个正 (i,j) 元. 但是，定理(6.2.16)说明，这种情况成立，当且仅当在 $\Gamma(A)$ 中存在从 P_i 到 P_j 的某条有向道路. 它等价于 $\Gamma(A)$ 是强连通的，而这又等价于 A 具有性质 SC. □

练习 证明推论(6.2.19)中涉及 $M(A)$ 的论断.

359 **6.2.20 推论** 在 $\Gamma(A)$ 中存在从 P_i 到 P_j 的道路，当且仅当 $[(I+|a|)^{n-1}]_{ij}\neq 0$.

练习 试用推论(6.2.19)给出关于性质 SC 的一个明确的计算检验法，它只需要对矩阵自乘大约 $\log_2(n-1)$ 次，而不是 $n-2$ 次. **提示**：考虑 $(I+|A|)^2$，这是平方，然后依此类推.

在结束这个论题以前，我们要引进性质 SC 的另外一个等价的描述. 它基于 $\Gamma(A)$ 的强连通性只是 $\Gamma(A)$ 的拓扑性质这一事实——它与 $\Gamma(A)$ 的诸结点所规定的下标毫无关系. 如果改变诸结点的下标顺序，图仍然保持其强连通性或非强连通性. 要注意的是，如果交换 A 的第 i 行和第 j 行以及第 i 列和第 j 列，这对 $\Gamma(A)$ 的影响是交换结点 P_i 和 P_j 的下标，反过来也

一样.

我们知道，转置矩阵 P 是方阵，它的所有元素为 0 或 1；在 A 的每一行和每一列恰好有一个 1. 显然，这样的矩阵是酉矩阵，因而是正交矩阵，所以 $P^T=P^{-1}$. 最简单的置换矩阵 P 对 i, j 的每个固定选择有 $p_{ij}=p_{ji}=1$，且所有其他非对角元为 0. 于是，相似 P^TAP 对 A 的作用是交换 A 的第 i 列和第 j 列以及交换 A 的第 i 行和第 j 行. A 的行和列的任一置换可以依次用这种形式的交换来得到. 并且任一置换矩阵是这样一些简单置换矩阵的有限乘积. 因此，如果 P 是置换矩阵，则通过适当置换 A 的若干行和列可由 A 得到相似 P^TAP. 重要的是要知道，是否可以求得 A 的行和列的某个置换把 A 变成下述特殊的分块形式.

6.2.21 定义 矩阵 A 称为可约的，指的是

(a) $n=1$ 且 $A=0$；或者

(b) $n\geqslant 2$，存在置换矩阵 $P\in M_n$，且存在适合 $1\leqslant r\leqslant n-1$ 的某个整数 r，使得

$$P^TAP=\begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

其中，$B\in M_r$，$D\in M_{n-r}$，$C\in M_{r,n-r}$，且 $0\in M_{n-r,r}$ 是零矩阵.

注意，我们并没要求子块 B，C 和 D 有非零元，而只要求通过一系列行的交换和列的交换可以在指定的位置上得到一个由 0 元组成的 $(n-r)\times r$ 子块. 如果 $|A|>0$，显然 A 不是可约的. 又如果 A 是可约，则它至少必须有 $(n-1)$ 个 0 元.

360

附注 假如想解线性方程组 $Ax=y$，且假定 A 是可约的. 于是，如果记 $\widetilde{A}=P^TAP=\begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$，则有 $Ax=P\widetilde{A}P^Tx=y$，或 $\widetilde{A}(P^Tx)=P^Ty$. 令 $P^Tx=\widetilde{x}=[z^T \vdots \zeta^T]^T$（未知）和 $P^Ty=\widetilde{y}=[w^T \vdots \omega^T]^T$（已知），其中，$z$，$w\in \mathbf{C}^r$，而 ζ，$\omega\in\mathbf{C}^{n-r}$. 因此，要解的方程组等价于 $\widetilde{A}\widetilde{x}=\widetilde{y}=\begin{bmatrix} B & C \\ 0 & D \end{bmatrix}\begin{bmatrix} z \\ \zeta \end{bmatrix}=\begin{bmatrix} w \\ \omega \end{bmatrix}$，即等价于

$$\begin{aligned} Bz+C\zeta&=w,\\ D\zeta&=\omega. \end{aligned}$$

如果首先对 ζ 来解 $D\zeta=\omega$，然后在第一个方程中采用 ζ 并且对 z 来解 $Bz=w-C\zeta$，那么就把原来的问题化简成两个较小的问题，一般说来，它们应该是比较容易解决的. 这番评注就是促成定义术语可约矩阵的理由.

6.2.22 定义 如果矩阵 $A\in M_n$ 不是可约的，就称 A 是不可约的.

6.2.23 定理 矩阵 $A\in M_n$ 是不可约的，当且仅当

$$(I+|A|)^{n-1}>0,$$

或等价地，如果 $[I+M(A)]^{n-1}>0$.

证明： 我们实际上要证明，A 是可约的，当且仅当 $(I+|A|)^{n-1}$ 至少有一个零元. 首先假定 A 是可约的，且对某个置换矩阵 P 有

$$A=P\begin{bmatrix} B & C \\ 0 & D \end{bmatrix}P^T=P\widetilde{A}P^T,$$

其中 B, C, 0 和 D 是定义(6.2.21)中的分块矩阵，注意到 P 的作用仅仅是交换行和列的顺序，所以 $|A|=|P\widetilde{A}P^T|=P|\widetilde{A}|P^T$；同时还注意到，$|\widetilde{A}|^2$，$|\widetilde{A}|^3$，…，$|\widetilde{A}|^{n-1}$和 $\widetilde{A}$ 一样，在其左下角都有相同的$(n-r)\times r$ 0 子块. 因此

$$(I+|A|)^{n-1}=(I+P|\widetilde{A}|P^T)^{n-1}=(P[I+|\widetilde{A}|]P^T)^{n-1}=P(I+|\widetilde{A}|)^{n-1}P^T$$

$$=P\left[I+(n-1)|\widetilde{A}|+\binom{n-1}{2}|\widetilde{A}|^2+\cdots+\binom{n-1}{n-1}|\widetilde{A}|^{n-1}\right]P^T,$$

且上述方括号中的所有项在其左下角都有$(n-r)\times r$ 0 子块. 于是，$(I+|A|)^{n-1}$是可约的，因而它的所有元不可能都是非零的.

361

反之，假定对某个 $p\neq q$，$(I+|A|)^{n-1}$的(p, q)元是 0. 于是得知在 $\Gamma(A)$中不存在从 P_p 到 P_q 的有向道路. 定义结点集

$$S_1=\{P_i:P_i=P_q\text{ 或在 }\Gamma(A)\text{ 中有一条从 }P_i\text{ 到 }P_q\text{ 的道路}\}$$

又设 S_2 包含 $\Gamma(A)$的所有不在 S_1 中的结点. 因 $S_1\cup S_2=\{P_1, \cdots, P_n\}$且 $P_q\in S_1\neq\varnothing$，所以 $S_2\neq\{P_1, \cdots, P_n\}$. 如果从 S_2 的某结点 P_i 到 S_1 的某结点 P_j 有一条道路，则(根据 S_1 的定义)存在从 P_i 到 P_q 的道路，因而 P_i 已在 S_1 中. 因此，从 S_2 的任一结点到 S_1 的任一结点不可能有任何道路. 现在我们重新标记结点，使 $S_1=\{\widetilde{P}_1, \cdots, \widetilde{P}_r\}$和 $S_2=\{\widetilde{P}_{r+1}, \cdots, \widetilde{P}_n\}$，注意到

$$\widetilde{A}=P^TAP=\begin{bmatrix}B & C\\ 0 & D\end{bmatrix},\quad B\in M_r,\quad 0\in M_{n-r,r},$$

因此 A 是可约的. 关于$[I+M(A)]^{n-1}>0$ 的证明不过是作同样的论述. □

我们作一下总结.

6.2.24 定理 设 $A\in M_n$. 则下列命题等价：

(a) A 是不可约矩阵；

(b) $(I+|A|)^{n-1}>0$；

(c) $[I+M(A)]^{n-1}>0$；

(d) $\Gamma(A)$是强连通的；

(e) A 具有性质 SC.

6.2.25 定义 设 $A\in M_n$. 称 A 是不可约对角占优矩阵，是指 A 适合以下条件：

(a) A 是不可约的；

(b) A 是对角占优的，即对所有 $i=1, \cdots, n$ 有 $|a_{ii}|\geqslant R_i'(A)$；

(c) 对 i 的至少一个值，有 $|a_{ii}|>R_i'(A)$.

练习 用例子说明，一个矩阵可能是不可约的和对角占优的，但不是不可约对角占优的.

362

采用现在的术语，可以重新表述"较好定理"(6.2.8)及其推论如下：

6.2.26 定理 设 $A\in M_n$ 是不可约的. 则仅当每个 Geršgorin 圆经过 λ 时，Geršgorin 区域 $G(A)$的边界点 λ 才可以为 A 的特征值.

6.2.27 推论(Taussky) 设 $A=[a_{ij}]\in M_n$ 是不可约对角占优矩阵. 那么，

(a) A 是可逆矩阵；

(b) 如果所有 $a_{ii}>0$，则对 A 的所有特征值 λ_i 有 $\mathrm{Re}(\lambda_i)>0$；

(c) 如果 A 是 Hermite 矩阵(或更一般地，A 只有实特征值)，又如果 A 的所有主对角元都是正的，则 A 的所有特征值都是正的.

6.2.28 推论 设 $A\in M_n$ 是不可约矩阵，且假定对 i 的至少一个值有

$$R_i=\sum_{j=1}^{n}|a_{ij}|<\|A\|_\infty,$$

即不是所有绝对行和等于极大绝对行和. 则 $\rho(A)<\|A\|_\infty$. 更一般地，如果 $p_1,\cdots,p_n>0$，如果

$$D=\mathrm{diag}(p_1,p_2,\cdots,p_n),$$

又如果对 i 的至少一个值有 $R_i(D^{-1}AD)<\|D^{-1}AD\|_\infty$，则 $\rho(A)<\|D^{-1}AD\|_\infty$.

证明：我们总有界 $\rho(A)\leqslant\|A\|_\infty$，而等式成立，当且仅当存在 A 的某个特征值 λ 适合 $|\lambda|=\|A\|_\infty$. 另一方面，根据定理(6.2.26)，每个 Geršgorin 圆必定经过 λ. 但是，某个 $R_i<\|A\|_\infty$ 的假设条件与这相矛盾. 把同样的证明应用于 $D^{-1}AD$ 便得出第二个论断. □

习题

1. 证明不可约矩阵不可能有 0 行和 0 列.

2. 用例子说明推论(6.2.28)中的不可约性假设是必不可少的.

3. 假定 $A=[a_{ij}]\in M_n$，λ 是 $|A|\equiv[|a_{ij}|]$ 的特征值，且存在所有 $x_i>0$ 的向量 $x=[x_i]\in\mathbf{R}^n$ 使得 $|A|x=\lambda x$. 设 $D=\mathrm{diag}(x_1,x_2,\cdots,x_n)$. 证明 $D^{-1}|A|D$ 的每个 Geršgorin 圆经过 λ，画出它的图形. 关于 $D^{-1}AD$ 的绝对行和，你能说些什么? 363

4. 在第 8 章中将证明，具有正元素的方阵总有正特征值和相应的正特征向量. 利用这一事实和上一个习题证明，只要 A 的所有元是非零的，就有

$$\rho(A)\leqslant\rho(|A|).$$

试用连续性证明，A 的所有元是非零的条件可以略去，因而对所有 $A\in M_n$ 有

$$\rho(A)\leqslant\rho(|A|).$$

5. 试用推论(6.2.28)证明，关于多项式

$$p(z)=z^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0,\quad a_0\neq 0$$

的根的 Cauchy 界(5.6.40)在不出现下述情形

$$|a_0|=|a_1|+1=|a_2|+1=\cdots=|a_{n-1}|+1$$

的假设下可稍许改进为

$$|z|<\max\{|a_0|,|a_1|+1,|a_2|+1,\cdots,|a_{n-1}|+1\}.$$

提示：证明，若 $a_0\neq 0$，则(5.6.39)中的友矩阵 $C(p)$ 是不可约的，对 Montel 界(5.6.41)，Carmichael 和 Mason 界(5.6.42)，Montel 界(5.6.43)，以及 Kojima 界(5.6.45)可作何改进?

进一步阅读 关于 Levy-Desplanques 定理的讨论以及许多有关的参考文献可参看 O. Taussky, "A Recurring Theorem on Determinants," *Amer. Math. Monthly* 56 (1949), 672-676.

6.3　扰动定理

设 $D=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)\in M_n$，设 $E=[e_{ij}]\in M_n$，并且考虑被扰动矩阵 $D+E$. 根据定理(6.1.1)，$D+E$ 的特征值在诸圆盘

$$\left\{z\in \mathbf{C}: |z-\lambda_i-e_{ii}|\leqslant R_i'(E)=\sum_{\substack{j=1\\ j\neq i}}^{n}|e_{ij}|\right\},\quad i=1,\cdots,n$$

364　中，它们又包含在诸圆盘

$$\left\{z\in \mathbf{C}: |z-\lambda_i|\leqslant R_i(E)=\sum_{j=1}^{n}|e_{ij}|\right\},\quad i=1,\cdots,n$$

中，因此，如果 $\hat{\lambda}$ 是 $D+E$ 的特征值，则存在 D 的某个特征值 λ_i，使得 $|\hat{\lambda}-\lambda_i|\leqslant\|E\|_\infty$. 遗憾的是，这个简单的估计不能推广到一般(非对角)的情形，但是，在矩阵是可对角化的情形，可以给出一个简单的界.

6.3.1 论断　设 $A\in M_n$ 是可对角化矩阵，且 $A=S\Lambda S^{-1}$ 和 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 设 $E\in M_n$. 如果 $\hat{\lambda}$ 是 $A+E$ 的一个特征值，则存在 Λ 的某个特征值 λ_i，使得

$$|\hat{\lambda}-\lambda_i|\leqslant\|S\|_\infty\|S^{-1}\|_\infty\|E\|_\infty=\kappa_\infty(S)\|E\|_\infty$$

其中 $\kappa_\infty(\cdot)$ 表示关于矩阵范数 $\|\cdot\|_\infty$ 的条件数.

证明：因为 $A+E$ 和 $S^{-1}(A+E)S=\Lambda+S^{-1}ES$ 有相同的特征值，又因为 Λ 是对角矩阵，所以，前述论证说明，存在某个 λ_i，使得 $|\hat{\lambda}-\lambda_i|\leqslant\|S^{-1}ES\|_\infty$. 因为 $\|\cdot\|_\infty$ 是矩阵范数，因此所述不等式成立. □

在技巧上稍作修改，便可把这个结果推广到不同于极大行和范数的其他矩阵范数. 矩阵范数的关键假设都被单调或绝对向量范数诱导的所有诱导矩阵范数满足；见(5.6.37).

6.3.2 定理　设 $A\in M_n$ 是可对角化矩阵，且 $A=S\Lambda S^{-1}$ 和 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 设 $E\in M_n$，又设 $\|\cdot\|$ 是这样的矩阵范数，使得 $\|D\|=\max_{1\leqslant i\leqslant n}|d_i|$ 对所有对角矩阵 $D=\mathrm{diag}(d_1, \cdots, d_n)\in M_n$ 成立. 如果 $\hat{\lambda}$ 是 $A+E$ 的一个特征值，则存在 A 的某个特征值 λ_i，使得

$$|\hat{\lambda}-\lambda_i|\leqslant\|S\|\,\|S^{-1}\|\,\|E\|=\kappa(S)\|E\|,\tag{6.3.3}$$

其中 $\kappa(\cdot)$ 是关于矩阵范数 $\|\cdot\|$ 的条件数.

证明：如同上一个结果那样，只要考虑 $S^{-1}(A+E)S=\Lambda+S^{-1}ES$ 的特征值就可以了. 如果 $\hat{\lambda}$ 是 $\Lambda+S^{-1}ES$ 的特征值，则 $\hat{\lambda}I-\Lambda-S^{-1}ES$ 是奇异矩阵. 如果 $\hat{\lambda}I-\Lambda$ 是奇异矩阵，则对每个 i 有 $\hat{\lambda}=\lambda_i$，因而界(6.3.3)显然被满足. 但是，假定 $\hat{\lambda}I-\Lambda$ 是非奇异矩阵. 在这种情形，矩阵

365　$$(\hat{\lambda}I-\Lambda)^{-1}(\hat{\lambda}I-\Lambda-S^{-1}ES)=I-(\hat{\lambda}I-\Lambda)^{-1}S^{-1}ES$$

是奇异矩阵，因而，根据(5.6.16)，一定有 $\|(\hat{\lambda}I-\Lambda)^{-1}S^{-1}ES\|\geqslant1$. 于是，因为假定矩阵范数 $\|\cdot\|$ 关于对角矩阵所具有的性质，有

$$1\leqslant\|(\hat{\lambda}I-\Lambda)^{-1}S^{-1}ES\|\leqslant\|S^{-1}ES\|\,\|(\hat{\lambda}I-\Lambda)^{-1}\|$$

$$=\|S^{-1}ES\|\max_{1\leqslant i\leqslant n}|\hat{\lambda}-\lambda_i|^{-1}=\frac{\|S^{-1}ES\|}{\min\limits_{1\leqslant i\leqslant n}|\hat{\lambda}-\lambda_i|},$$

因此

$$\min_{1\le i\le n} |\hat{\lambda}-\lambda_i| \leqslant \|S^{-1}ES\| \leqslant \|S^{-1}\| \|S\| \|E\| = \kappa(S)\|E\|. \qquad \Box$$

练习 证明，定理(6.3.2)假定矩阵范数在对角矩阵上所具有的性质，下列各个范数都满足：$\|\cdot\|_2$，$\|\cdot\|_\infty$，$\|\cdot\|_1$. 至少给出另一个矩阵范数，它也满足上述假设.

练习 给出一个矩阵范数，它不满足定理(6.3.2)的假设.

练习 证明$\|U\|_2=1$对任何酉矩阵成立.

虽然在(5.8)中曾出现过条件数$\kappa(\cdot)$，它是由于线性方程组解的误差界引起的，但是我们看到，现在在(6.3.3)中也出现了条件数，不过它是在计算可对角化矩阵的特征值中作为误差比

$$\frac{|\hat{\lambda}-\lambda_i|}{\|E\|} \leqslant \kappa(S)$$

的上界而出现的. 如果$\kappa(S)$较小(接近1)，则小的数据扰动可以使特征值产生扰动，不过特征值的变化范围以数据变化的相同数量级为界. 但是，如果$\kappa(S)$很大，则小的数据扰动可能引起特征值的比较大的变化.

与(5.8)中关于线性方程组解的情况不同的是，在这里$\kappa(A)$没有多大意义，但是$\kappa(S)$是很重要的，其中，$A=SAS^{-1}$，而S是以A的特征向量为列的矩阵. 关于谱范数的条件数具有$\kappa(S)=\cot(\theta/2)$的几何解释，其中θ是当x和y取遍所有可能的非零正交向量时Sx与Sy之间的最小夹角[见例(7.1.26)]. 因此，它不依赖A的条件数. 如果A的一对线性无关的特征向量接近平行，那么S的两列(例如，第p列和第q列，$p\neq q$)可能接近平行，因而，即使单位基向量e_p和e_q是正交的，Se_p与Se_q的夹角仍可能很小. 在这种情形，谱条件数$\kappa(S)$将很大，而确定A的特征值问题就可能是病态的. 366

但是，如果S是酉(或接近于酉)矩阵，则S将把一对正交向量变成正交(或将近于正交)向量，并且S的谱条件数将会很小(实际上，如果S是酉矩阵，它等于1). 在这种情形，确定A的特征值的问题一定是良态的. 当然，一个矩阵(恰好)可以酉对角化，当且仅当它是正规矩阵. 所以，(6.3.2)给出了关于整个正规矩阵(特别是Hermite矩阵和实对称矩阵)类的扰动定理，它与原来的关于对角矩阵的论断具有同样简单的形式. 正规矩阵关于特征值的计算是优态的.

6.3.4 推论 设$A\in M_n$是具有特征值λ_1，…，λ_n的正规矩阵，又设$E\in M_n$. 如果$\hat{\lambda}$是$A+E$的特征值，则存在A的某个特征值λ_i，使得$|\hat{\lambda}-\lambda_i|\leqslant\|E\|_2$.

注意，扰动矩阵E和被扰动矩阵$A+E$都不一定是正规矩阵. 推论(6.3.4)常常应用于实对称矩阵A的情形.

练习 给出推论(6.3.4)的详细证明.

练习 如果知道A和E都是Hermite矩阵，则可以利用Weyl定理(4.3.1)给出比(6.3.4)中的界更好的界. 如果A，$E\in M_n$是Hermite矩阵，又如果$\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$是$A$的有序特征值，$\hat{\lambda}_1\leqslant\hat{\lambda}_2\leqslant\cdots\leqslant\hat{\lambda}_n$是$A+E$的有序特征值，且$\lambda_1(E)\leqslant\cdots\leqslant\lambda_n(E)$是$E$的有序特征值，试用不等式(4.3.2)证明，对所有$k=1,2,\cdots,n$，

$$\lambda_1(E) \leqslant \hat{\lambda}_k - \lambda_k \leqslant \lambda_n(E).$$

且

$$|\hat{\lambda}_k - \lambda_k| \leqslant \rho(E) = \|E\|_2.$$

说明为什么这个界比(6.3.4)好. 如果已知 E 的所有特征值是非负的，这提供了什么信息？

367 在数值应用中，原矩阵 A 和扰动矩阵 E 都是实对称的情形并不少见. 在这种情形以及在 A 和 $A+E$ 都是正规矩阵的更一般的情形，关于扰动，存在一个对所有特征值都适用的涉及全局的界.

6.3.5 定理(Hoffman 和 Wielandt) 设 A, $E\in M_n$. 假定 A 和 $A+E$ 都是正规矩阵，设 $\{\lambda_1, \cdots, \lambda_n\}$ 是按某个顺序给定的 A 的特征值，$\{\hat{\lambda}_1, \cdots, \hat{\lambda}_n\}$ 是按某个顺序给定的 $A+E$ 的特征值. 则存在整数 1, 2, …, n 的一个排列 $\sigma(i)$ 使得

$$\left[\sum_{i=1}^{n} |\hat{\lambda}_{\sigma(i)} - \lambda_i|^2\right]^{1/2} \leqslant \|E\|_2. \tag{6.3.6}$$

证明：设 $\Lambda=\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$, $\hat{\Lambda}=\operatorname{diag}(\hat{\lambda}_1, \cdots, \hat{\lambda}_n)$, 设 $V\in M_n$ 是使 $A=V\Lambda V^*$ 的酉矩阵，而 $W\in M_n$ 是使 $A+E=W\hat{\Lambda}W^*$ 的酉矩阵. 于是，因为 Frobenius 范数是酉不变的，有

$$\begin{aligned}
\|E\|_2^2 &= \|(A+E)-A\|_2^2 \\
&= \|W\hat{\Lambda}W^* - V\Lambda V^*\|_2^2 \\
&= \|V^*W\hat{\Lambda}W^*V - \Lambda\|_2^2 \\
&= \|Z\hat{\Lambda}Z^* - \Lambda\|_2^2 \\
&= \operatorname{tr}(Z\hat{\Lambda}Z^* - \Lambda)(Z\hat{\Lambda}Z^* - \Lambda)^* \\
&= \operatorname{tr}(\hat{\Lambda}\hat{\Lambda}^* + \Lambda\Lambda^*) - \operatorname{tr}(Z\hat{\Lambda}Z^*\Lambda^* + \Lambda Z\hat{\Lambda}^*Z^*) \\
&= \sum_{i=1}^{n}(|\hat{\lambda}_i|^2 + |\lambda_i|^2) - 2\operatorname{Re}\operatorname{tr}(Z\hat{\Lambda}Z^*\Lambda^*),
\end{aligned}$$

其中已令 $Z\equiv V^*W$. 这个表示式说明

$$\|E\|_2^2 \geqslant \sum_{i=1}^{n}(|\hat{\lambda}_i|^2 + |\lambda_i|^2) - 2\max\{\operatorname{Re}\operatorname{tr}(U\hat{\Lambda}U^*\Lambda^*) : U \text{ 是酉矩阵}\}. \tag{6.3.7}$$

我们要证明，这个下界的精确值就是所确定的界(6.3.6). 如果 $U\equiv[u_{ij}]\in M_n$，则容易算出

$$\operatorname{Re}\operatorname{tr}(U\hat{\Lambda}U^*\Lambda^*) = \sum_{i,j=1}^{n} |u_{ij}|^2 \operatorname{Re}(\bar{\lambda}_i\hat{\lambda}_j),$$

而我们感兴趣的是，当 U 取遍所有 $n\times n$ 酉矩阵组成的紧集时，这个表示式的极大值. 如果令 $c_{ij}=|u_{ij}|^2$，且设 $C\equiv[c_{ij}]$，则矩阵 $C\in M_n$ 就是具有非负元的矩阵，且它的所有行和与所有列和恰好都是+1(因为 $UU^*=U^*U=I$). 因而，只要 U 是酉矩阵，C 就是双随机矩阵. 又如果修改我们的极值问题，允许取所有双随机矩阵，则我们将在一个知道其结构的紧凸集上更有利地考虑这个极值问题，在这个较大的区域上的极大值当然可能更大一些：368

$$\begin{aligned}
\max\{\operatorname{Re}\operatorname{tr}(U\hat{\Lambda}U^*\Lambda^*) : U \text{ 是酉矩阵}\} &= \max\left\{\sum_{i,j=1}^{n} |u_{ij}|^2 \operatorname{Re}(\bar{\lambda}_i\hat{\lambda}_j) : U \text{ 是酉矩阵}\right\} \\
&\leqslant \max\left\{\sum_{i,j=1}^{n} c_{ij}\operatorname{Re}(\bar{\lambda}_i\hat{\lambda}_j) : C \text{ 是双随机矩阵}\right\}
\end{aligned}$$

但是，这个求极大值的函数是紧凸集上的线性函数，所以在该凸集的一个端点上取得极大值（见附录 B 并且注意到线性函数是凸函数）．根据 Birkhoff 定理(8.7.1)，双随机矩阵集合的各端点都是置换矩阵，因而存在置换矩阵 $P\in M_n$，使得

$$\max\{\sum_{i,j=1}^{n}c_{ij}\operatorname{Re}(\bar{\lambda}_i\hat{\lambda}_j), C\text{是双随机矩阵}\}=\operatorname{Re}\operatorname{tr}(P\hat{\Lambda}P^T\Lambda^*).$$

因为置换矩阵是酉矩阵，所以也有

$$\max\{\operatorname{Re}\operatorname{tr}(U\hat{\Lambda}U^*\Lambda^*); U\text{是酉矩阵}\}=\operatorname{Re}\operatorname{tr}(P\hat{\Lambda}P^T\Lambda^*).$$

如果对 $i=1, 2, \cdots, n$ 有 $Pe_i=e_{\sigma(i)}$，则

$$\operatorname{Re}\operatorname{tr}(P\hat{\Lambda}P^T\Lambda^*)=\sum_{i=1}^{n}\operatorname{Re}(\hat{\lambda}_{\sigma(i)}\bar{\lambda}_i),$$

而(6.3.7)说明

$$\begin{aligned}\|E\|_2^2&\geqslant\sum_{i=1}^{n}[|\hat{\lambda}_{\sigma(i)}|^2+|\lambda_i|^2-2\operatorname{Re}(\hat{\lambda}_{\sigma(i)}\bar{\lambda}_i)]\\&=\sum_{i=1}^{n}|\hat{\lambda}_{\sigma(i)}-\lambda_i|^2.\end{aligned}$$

□

定理(6.3.5)说明，对正规矩阵的特征值集合，存在很强的全局稳定性，但是它没有说明究竟是特征值的哪个排列将满足所述不等式．并不是特征值的每个排列都满足这个不等式，实际上，至少存在特征值的一个排列，可使(6.3.6)的不等式反向（见本节末习题 7）．然而，在 Hermite 矩阵这一重要特殊情形，特征值的自然顺序能满足(6.3.6)中的不等式．

369

6.3.8 推论 设 $A, E\in M_n$，假定 A 是 Hermite 矩阵，$A+E$ 是正规矩阵，设$\{\lambda_1, \cdots, \lambda_n\}$是 A 的特征值，且排成递增顺序($\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$)，又设$\{\hat{\lambda}_1, \cdots, \hat{\lambda}_n\}$是 $A+E$ 的特征值，使其有顺序 $\operatorname{Re}\hat{\lambda}_1\leqslant\operatorname{Re}\hat{\lambda}_2\leqslant\cdots\leqslant\operatorname{Re}\hat{\lambda}_n$．则

$$\left[\sum_{i=1}^{n}|\hat{\lambda}_i-\lambda_i|^2\right]^{1/2}\leqslant\|E\|_2.$$

证明：根据定理(6.3.5)，存在 $A+E$ 的特征值的给定顺序（递增实部）的某个排列 σ，使得

$$\left[\sum_{i=1}^{n}|\hat{\lambda}_{\sigma(i)}-\lambda_i|^2\right]^{1/2}\leqslant\|E\|_2. \tag{6.3.9}$$

如果 $A+E$ 的特征值在表$\hat{\lambda}_{\sigma(1)}, \cdots, \hat{\lambda}_{\sigma(n)}$中已使它们的实部成递增顺序，那就没有什么可证的了．否则，在上述表中有两个相邻的特征值，其实部不按递增顺序排列，例如，对某个适合 $1\leqslant k<n$ 的 k，

$$\operatorname{Re}\hat{\lambda}_{\sigma(k)}>\operatorname{Re}\hat{\lambda}_{\sigma(k+1)}.$$

但是，因为

$$\begin{aligned}|\hat{\lambda}_{\sigma(k)}-\lambda_k|^2+|\hat{\lambda}_{\sigma(k+1)}-\lambda_{k+1}|^2=&|\hat{\lambda}_{\sigma(k+1)}-\lambda_k|^2+|\hat{\lambda}_{\sigma(k)}-\lambda_{k+1}|^2\\&+2(\lambda_k-\lambda_{k+1})(\operatorname{Re}\hat{\lambda}_{\sigma(k+1)}-\operatorname{Re}\hat{\lambda}_{\sigma(k)}),\end{aligned}$$

又因为根据假定，$\lambda_k-\lambda_{k+1}\leqslant0$，所以得知

$$|\hat{\lambda}_{\sigma(k)}-\lambda_k|^2+|\hat{\lambda}_{\sigma(k+1)}-\lambda_{k+1}|^2\geqslant|\hat{\lambda}_{\sigma(k+1)}-\lambda_k|^2+|\hat{\lambda}_{\sigma(k)}-\lambda_{k+1}|^2.$$

因此，可以交换两个特征值$\hat{\lambda}_{\sigma(k)}$和$\hat{\lambda}_{\sigma(k+1)}$且不增加平方差之和．通过有限次这样的交换，特征

值表 $\hat{\lambda}_{\sigma(1)}$，…，$\hat{\lambda}_{\sigma(n)}$ 可以变换成表 $\hat{\lambda}_1$，$\hat{\lambda}_2$，…，$\hat{\lambda}_n$，使其实部是递增的，且所确定的界成立.

□

实际上，这个推论经常应用于 A 和 $A+E$ 都是 Hermite 矩阵或者甚至于是实对称矩阵的情形.

练习 证明，如果 A，$B\in M_n$ 是 Hermite 矩阵又如果它们的特征值都按递增顺序或都按递减顺序排列，则

$$\left(\sum_{i=1}^{n}[\lambda_i(A)-\lambda_i(B)]^2\right)^{1/2}\leqslant \|A-B\|_2.$$

练习 说明，如果 A 和 $B=A+E$ 不都是正规矩阵，则定理(6.3.5)中的结论不一定成立.

[370] **提示**：设 $A=\begin{bmatrix}0&0\\0&4\end{bmatrix}$，$B=\begin{bmatrix}-1&-1\\1&1\end{bmatrix}$，然后说明，对特征值的任一顺序，

$$\sum_{i=1}^{2}[\lambda_i(A)-\lambda_i(B)]^2=16.$$

如果 A 不可对角化，就不知道有像定理(6.3.2)中那样简单的界. 但是有可能导出一个简明的公式，它说明，当矩阵的元素产生扰动时，矩阵的代数单重特征值(代数重数等于1)如何变化. 首先给出一个引理，说明相应于一个单重特征值的左特征向量与右特征向量具有非正交性.

6.3.10 引理 如果 $A\in M_n$，且 λ 是 A 的代数单重特征值，又如果 x 和 y 分别是相应于 A 的特征值 λ 的右特征向量和左特征向量，则 $y^*x\neq 0$.

证明：如果 $Ax=\lambda x$，$x\neq 0$，则可以采用在 Schur 三角化定理(2.3.1)的证明中所使用过的方法，构造其第1列是 $x/\|x\|_2$ 的酉矩阵 U，使得

$$U^*AU=\begin{bmatrix}\lambda & *\\ 0 & B\end{bmatrix},B\in M_{n-1}.$$

因为 λ 是 A 的单重特征值，所以它不可能是 B 的特征值. 单位基向量 e_1 是 U^*AU 的属于 λ 的特征向量. 现在考虑

$$(U^*AU)^*=U^*A^*U=\begin{bmatrix}\bar{\lambda} & 0\\ * & B^*\end{bmatrix},$$

且假定 $U^*A^*Uz=\bar{\lambda}z$，其中 $z\neq 0$. 如果 $z^*=[0|\xi^*]$，则 $\xi\neq 0$ 且 ξ 是 B^* 的属于 $\bar{\lambda}$ 的特征向量. 另一方面，λ 就应是 B 的特征值，这与假设相矛盾. 由此得知，z 不能以零作为第一个分量，即 $z^*e_1\neq 0$. 但是 $(Uz)^*(Ue_1)=z^*e_1\neq 0$，且向量 Uz 和 Ue_1 是 A 的属于 λ 的左特征向量和右特征向量. 因为根据假定，A 的关于 λ 的左特征空间和右特征空间都是一维的，所以对某个 $\alpha\neq 0$，一定有 $y=\alpha Uz$. 但是 $x=\|x\|_2Ue_1$，因此 $y^*x\neq 0$ 一定成立. □

练习 考虑 $A=\begin{bmatrix}1&1\\0&1\end{bmatrix}$，说明，如果从假设条件中略去"代数单重"，则引理不成立.

现在假定 λ 是 A 的代数单重特征值. 于是 A 有唯一(除了差一个纯量因子 α 以外，$|\alpha|=1$)确定的正规化右 λ 特征向量 x 和唯一确定的左 λ 特征向量 y，它可以通过关系 $y^*x=1$ 正

[371]

规化．如果考虑可微的参数族 $A(t)$，使得 $A(0)=A$[例如，$A(t)=A+tE$，E 为固定的扰动矩阵]，则对所有充分小的 t，存在 $A(t)$的唯一确定的单重特征值$\lambda(t)$，使得$\lambda(0)=\lambda$．同时存在一个右$\lambda(t)$的特征向量 $x(t)$，它由条件 $x^*(t)x(t)\equiv 1$ 唯一确定(如前，差一个因子 α)还存在一个左$\lambda(t)$特征向量 $y(t)$，它由条件 $y^*(t)x(t)\equiv 1$ 唯一确定．

如果微分这后一个正规化条件，则得到恒等式

$$y'^*(t)x(t)+y^*(t)x'(t)=0. \tag{6.3.11}$$

因为 $A(t)x(t)=\lambda(t)x(t)$对所有小的 t 成立，所以也有恒等式 $y^*(t)A(t)x(t)=\lambda(t)y^*(t)x(t)=\lambda(t)$．如果微分这个恒等式，便得出

$$\lambda'(t)=y'^*(t)A(t)x(t)+y^*(t)A'(t)x(t)+y^*(t)A(t)x'(t).$$

但是，因为 $A(t)x(t)=\lambda(t)x(t)$和 $y^*(t)A(t)=\lambda(t)y^*(t)$，这就变成

$$\lambda'(t)=\lambda(t)\{y'^*(t)x(t)+y^*(t)x'(t)\}+y^*(t)A'(t)x(t)=y^*(t)A'(t)x(t).$$

我们已经用到了恒等式(6.3.11)．在 $t=0$，这在正规化条件 $x^*x=1$ 和 $y^*x=1$ 下，正好是恒等式$\lambda'(0)=y^*A'(0)x$．如果 x 和 y 是右λ 特征向量和左λ 特征向量，它们不一定用上述方式正规化．可以用 $x/(x^*x)^{1/2}$代替 x，用$(x^*x)^{1/2}y/x^*y$ 代替 y 来得到一般的恒等式$\lambda'(0)y^*x=y^*A'(0)x$．至此已经证明了关于矩阵 A 的下述结果，它不要求 A 一定可对角化．

6.3.12 定理 设 $A(t)\in M_n$ 在 $t=0$ 可微．假定λ是$A(0)$的代数单重特征值，并且假定对小的 t，$\lambda(t)$是 $A(t)$的特征值，且使$\lambda(0)=\lambda$．设 x 是 A 的右λ 特征向量，y 是 A 的左λ 特征向量，则

$$\lambda'(0)=\frac{y^*A'(0)x}{y^*x}.$$

练习 设 $A(t)=A+tE$，其中 E 是固定的扰动矩阵，(在定理(6.3.12)的假定下)证明，在 $t=0$，

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t}=\frac{y^*Ex}{y^*x}.$$

练习 在定理(6.3.12)假定下．证明，对任意 i，j，

$$\frac{\partial\lambda}{\partial a_{ij}}=\frac{\bar{y}_ix_j}{y^*x}.$$

372

这个公式说明，相对于 A 的任一元素的变化，λ如何变化．**提示**：设 $E=E_{ij}$，它是 $n\times n$ 矩阵，且其仅有的非零元在 i，j 位置上．

练习 考察矩阵 $A=\begin{bmatrix}1 & 1\\ 0 & 1+\varepsilon\end{bmatrix}$，如果$\varepsilon\neq 0$，特征值$\lambda=1$ 就是单重的．对所有四对 i，j，直接算出 $\partial\lambda/\partial a_{ij}$．当$\varepsilon\to 0$ 时，这些变化具有什么性质？由此可知，如果 x 和 y 接近正交，那么，对于 A 的某些扰动，特征值λ可能是很灵敏的．

与特征值的情形形成对比，仅仅因为矩阵元素的小扰动，即使是对角矩阵的特征向量，也可能会有很大的变化．例如，如果 $A=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$，且 $E=\begin{bmatrix}\varepsilon & \delta\\ 0 & 0\end{bmatrix}$，其中$\varepsilon$，$\delta\neq 0$，则 $A+E$ 的特征值是$\lambda=1$ 和 $1+\varepsilon$，而相应的正规化特征向量是

$$\frac{1}{(\varepsilon^2+\delta^2)^{1/2}}\begin{bmatrix}\delta\\ \varepsilon\end{bmatrix} \quad 和 \quad \begin{bmatrix}1\\ 0\end{bmatrix}.$$

对于任意小的 ε 和 δ，通过适当选择 ε 与 δ 之比，可以把第一个特征向量选定为任意方向上的点已解.

如果令 $\varepsilon=0$，则对任意 $\delta\neq0$，被扰动矩阵 $A+E=\begin{bmatrix}1 & \delta\\ 0 & 1\end{bmatrix}$只有一个无关的特征向量，而 A 却有两个无关的特征向量.

迄今我们的所有估计是关于矩阵的扰动所引起的特征值误差的先验界；它们不涉及所计算的特征值或特征向量，或由它们导出的任何量. 假定"近似特征向量"$\hat{x}\neq0$ 和"近似特征值"$\hat{\lambda}$ 已经用某种方式求出来，可能不是 $A\hat{x}$ 恰好等于 $\hat{\lambda}\hat{x}$ 的情形，但是，当 A 可对角化时，可以利用剩余向量 $r=A\hat{x}-\hat{\lambda}\hat{x}$ 得到一个估计，确定 $\hat{\lambda}$ 是如何接近 A 的一个特征值.

记 $A=S\Lambda S^{-1}$，且假定 $\hat{\lambda}$ 不恰好等于 A 的任何特征值. 于是

$$r = A\hat{x}-\hat{\lambda}\hat{x} = S(\Lambda-\hat{\lambda}I)S^{-1}\hat{x},$$

所以 $\hat{x}=S(\Lambda-\hat{\lambda}I)^{-1}S^{-1}r$. 因此，

$$\begin{aligned}\|\hat{x}\| &= \|S(\Lambda-\hat{\lambda}I)^{-1}S^{-1}r\| \leqslant |\!|\!|S(\Lambda-\hat{\lambda}I)^{-1}S^{-1}|\!|\!|\,\|r\|\\ &\leqslant |\!|\!|S|\!|\!|\,|\!|\!|S^{-1}|\!|\!|\,|\!|\!|(\Lambda-\hat{\lambda}I)^{-1}|\!|\!|\,\|r\| = \kappa(S)|\!|\!|(\Lambda-\hat{\lambda}I)^{-1}|\!|\!|\,\|r\|\\ &= \kappa(S)\left(\min_{1\leqslant i\leqslant n}|\lambda_i-\hat{\lambda}|\right)^{-1}\|r\|,\end{aligned}$$

[373]

从而

$$\|\hat{x}\|\min_{1\leqslant i\leqslant n}|\lambda_i-\hat{\lambda}|\leqslant\kappa(S)\|r\|.$$

显然，即使当某个 $\lambda_i=\hat{\lambda}$ 时，后一个不等式仍然成立. 对于上述论证，已经假定

$$\left.\begin{array}{l}\text{(a) } \|\cdot\| \text{ 是 } \mathbf{C}^n \text{ 上的向量范数；}\\ \text{(b) } M_n \text{ 上的矩阵范数} |\!|\!|\cdot|\!|\!| \text{与} \|\cdot\| \text{ 相容；}\\ \text{(c) } |\!|\!|D|\!|\!|=\max\limits_{1\leqslant i\leqslant n}|d_i|\text{，如果 } D=\mathrm{diag}(d_1,\cdots,d_n)\in M_n\text{；}\end{array}\right\}\qquad(6.3.13)$$

且条件数 $\kappa(S)$ 是用矩阵范数$|\!|\!|\cdot|\!|\!|$来计算的. 如果 A 是正规矩阵，则 S 可以取为酉矩阵，又如果采用 l_2 向量范数和谱矩阵范数，则有 $\kappa(S)=1$. 条件(c)等价于要求矩阵范数$|\!|\!|\cdot|\!|\!|$是由单调向量范数诱导的[定理(5.6.37)]. 因此，如果 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上的单调向量范数，且$|\!|\!|\cdot|\!|\!|$是由 $\|\cdot\|$ 诱导的 M_n 上矩阵范数，则(6.3.13)的所有条件都被满足. 这样，就证明了关于后验界的结果，它与定理(6.3.2)和推论(6.3.4)具有相同的形式.

6.3.14 定理 设 $A\in M_n$ 是可对角化矩阵，且 $A=S\Lambda S^{-1}$ 和 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$. 设 $\mathbf{C}^n$ 上的向量范数 $\|\cdot\|$ 和 M_n 上的矩阵范数$|\!|\!|\cdot|\!|\!|$适合条件(6.3.13)，设 $\hat{x}\in\mathbf{C}^n$ 是给定的非零向量，$\hat{\lambda}$ 是给定的复数，且 $r=A\hat{x}-\hat{\lambda}\hat{x}$. 则存在 A 的某个特征值 λ_i，使得

$$|\hat{\lambda}-\lambda_i|\leqslant|\!|\!|S|\!|\!|\,|\!|\!|S^{-1}|\!|\!|\frac{\|r\|}{\|\hat{x}\|}=\kappa(S)\frac{\|r\|}{\|\hat{x}\|}.\qquad(6.3.15)$$

如果 A 是正规矩阵，则存在 A 的某个特征值 λ_i，使得

$$|\hat{\lambda}-\lambda_i| \leqslant \frac{\|r\|_2}{\|\hat{x}\|_2}. \tag{6.3.16}$$

这后一个结果可能不同于线性方程组的解的相对误差的后验界的相应结果. 如果线性方程组的系数矩阵是病态的，(5.8.11)说明，小的剩余向量并不蕴涵解的小的相对误差. 然而，(6.3.16)却说明，如果 A 是正规矩阵(实际上，A 通常是 Hermite 矩阵或实对称矩阵)，又如果近似特征向量——特征值偶有小的剩余向量，则可以保证特征值的绝对误差是小的；没有条件数出现在界中.

[374]

这个关于特征值的惬意结果没有关于特征向量的类似的惬意结果相匹配. 即使对于实对称矩阵，一个小的剩余向量并不保证近似特征向量接近于一个特征向量. 例如，考虑 $A=\begin{bmatrix}1 & \epsilon\\ \epsilon & 1\end{bmatrix}$，$\epsilon>0$. 如果取 $\hat{\lambda}=1$ 和 $\hat{x}=[1,\ 0]^T$，则剩余向量是 $r=[0,\ \epsilon]^T$. 对所有 $\epsilon>0$，A 的特征向量是 $[1,\ 1]^T$ 和 $[1,\ -1]^T$，而不论 ϵ 如何小，$\hat{x}$ 不近似平行于这两个向量中的任何一个.

练习 证明上述例子中 A 的特征值是 $1+\epsilon$ 和 $1-\epsilon$，并且验证在这种情形下的界(6.3.16).

习题

1. 如果 λ，μ 是 A 的特征值，且 $\lambda\neq\mu$，证明 A 的相应于 μ 的任一左特征向量与 A 的相应于 λ 的任一右特征向量正交.

2. 在 A 有互不相同的特征值的假定下，利用上一个习题给出引理(6.3.10)的另一个证明.

3. 考察

$$A_\epsilon=\begin{bmatrix}0 & 1\\ \epsilon & 0\end{bmatrix}\in M_2,\quad A_0=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix},$$

其中 $\epsilon\geqslant 0$，试验证本节第一段最后一句话中所表现出的遗憾. 证明，对 $\epsilon>0$，A_ϵ 可对角化，且 A_ϵ 的一个特征值与 A_0 的每一个特征值间的极小距离是$\sqrt{\epsilon}$. 记 $A_\epsilon=A_0+E$. 并证明

$$\frac{|\hat{\lambda}-\lambda_i|}{\|E\|}\geqslant O(\epsilon^{-1/2})\to\infty,\quad 若\ \epsilon\to 0$$

因此，一般说来，没有形如 $|\hat{\lambda}-\lambda_i|\leqslant\|E\|$ 的界可能是正确的. 然后计算在这种情形下定理(6.3.2)的界，并且说明会出现什么情形.

4. 考察多项式 $p(x)=(x-x_0)^2$，它在 x_0 有重根，即 $p(x_0)=p'(x_0)=0$，但 $p''(x_0)=0$. 证明，对较小的 $\epsilon>0$，$p(x)-\epsilon$ 有形如 $x_0\pm\epsilon^{1/2}$ 与 x_0 接近的两个根. 因此，一个多项式的各个系数中阶 ϵ 的一个改变就可以使它的各个根按阶$\sqrt{\epsilon}$的一个相当量发生改变. 对于一个多项式，一个零点的扰动与各系数的扰动之比可能是无界的.

[375]

5. 考虑界(6.3.4)，它是说，对于 Hermite 矩阵(或更一般地，正规矩阵)，特征值的扰动与矩阵诸元素的扰动之比是有界的. 因为矩阵的各特征值正好是其特征多项式的各个根. 解释为何这个合意情形与习题 4 中的结论是一致的. 历史的经验是，按传统习惯去构造特征多项式然后求它的各根的方法来计算 Hermite 矩阵(或任何其他矩阵)的各特征值是不可取的. 这有可

能把原本良态的问题转为病态问题!

6. 考虑 Givens 给的例子，设 $A=I$ 是 2×2 实对称矩阵，而

$$E(\varepsilon)=\begin{bmatrix}\varepsilon\cos(2/\varepsilon) & \varepsilon\sin(2/\varepsilon)\\ \varepsilon\sin(2/\varepsilon) & -\varepsilon\cos(2/\varepsilon)\end{bmatrix},\quad \varepsilon>0$$

是一个实对称扰动矩阵，且 $E(0)\equiv\lim_{\varepsilon\to0}E(\varepsilon)=0$. 证明，$A+E(\varepsilon)$的特征值是 $1+\varepsilon$ 和 $1-\varepsilon$，而 $A+E(\varepsilon)$相应的(唯一确定到相差一个符号)的正规化特征向量分别是$[\cos(1/\varepsilon),\ \sin(1/\varepsilon)]^T$ 和 $[\sin(1/\varepsilon),\ -\cos(1/\varepsilon)]^T$，其中 $\varepsilon>0$. 证明，当 $\varepsilon\to0$ 时，每个特征向量可随意地指向任一给定的方向. 因此，即使我们仅考虑实对称矩阵，如果它的特征值与其他特征值没有明显的区别，则有个别特征向量可以快速地变化.

7. 试用定理(6.3.5)的证法证明，(在该定理的假设条件下)存在整数 1，2，…，n 的一个排列 τ，使得

$$\left(\sum_{i=1}^{n}|\hat{\lambda}_{\tau(i)}-\lambda_i|^2\right)^{1/2}\geqslant\|E\|_2.$$

提示：考虑 $\min\left\{\sum_{i,j=1}^{n}c_{ij}\mathrm{Re}(\hat{\lambda}_i\bar{\lambda}_j):C=[c_{ij}]\text{ 是双随机矩阵}\right\}$.

8. 设 $A\in M_n$ 是给定的正规矩阵，其特征值集为$\{\lambda_i(A)\}$，又设 $r>0$ 是给定的，且定义

$$S(A,r)\equiv\{B\in M_n:B\text{ 是正规矩阵,且 }\|B-A\|_2\leqslant r\}$$

证明$\{\hat{\lambda}_1,\ \cdots,\ \hat{\lambda}_n\}$是矩阵 $B\in S(A,\ r)$的特征值集合当且仅当

$$\min\left\{\sum_{i=1}^{n}|\lambda_i(A)-\hat{\lambda}_{\sigma(i)}|^2:\sigma\text{ 是 }1,\cdots,n\text{ 的一个排列}\right\}\leqslant r^2.$$

这给出了一个给定的正规矩阵的邻域中的诸正规矩阵的可能特征值集合的完整特征. **提示**：必要性用定理(6.3.5). 关于充分性，设 $A=U\Lambda U^*$，其中 $\Lambda=\mathrm{diag}(\lambda_1(A),\ \cdots,\ \lambda_n(A))$,然后定义 $B=U\hat{\Lambda}U^*$，其中 $\hat{\Lambda}=\mathrm{diag}(\hat{\lambda}_1,\ \cdots,\ \hat{\lambda}_n)$.

9. 在定理(6.3.5)的证明中，用到了如下事实：如果 $U=[u_{ij}]\in M_n$ 是酉矩阵，则 $A=[|u_{ij}|^2]$是双随机矩阵. 说明不是每个双随机矩阵可以用这种方式由酉矩阵产生. **提示**：考虑例子

$$\frac{1}{2}\begin{bmatrix}1&1&0\\1&0&1\\0&1&1\end{bmatrix}.$$

10. 假定 $A\in M_3$ 是给定的 Hermite 矩阵，且假定用某种方式已求得一个酉矩阵 U，使得

$$UAU^*=\begin{bmatrix}3.05&-0.06&0.02\\-0.06&-6.91&0.07\\0.02&0.07&8.44\end{bmatrix}$$

对于 A 的特征值，给出你所能做到的最佳估计.

11. 不能指望非正规矩阵有形如(6.3.4)的界. 考虑 A、$E\in M_n$，其中，

$$A=\begin{bmatrix}0 & a & & & \\ & & 0 & & 0\\ & & & \ddots & \ddots\\ 0 & & & & 0\\ & & & & 0\end{bmatrix},$$

$$E=\begin{bmatrix}0 & 0 & & & \\ & & \varepsilon & & 0\\ \vdots & & & \varepsilon & \\ & 0 & & \ddots & \ddots\\ 0 & & & & \varepsilon\\ \varepsilon & 0 & & \cdots & 0\end{bmatrix},\quad a,\varepsilon\geqslant 0.$$

证明 A 的所有特征值都是 0，而 $A+E$ 的特征值是关于$\sqrt[n]{a\varepsilon^{n-1}}$的 n 个不同的值. 无论 $\varepsilon>0$ 为何值，只要适当地选取 a，就可使 $A+E$ 的所有特征值任意大. 当 A 是正规矩阵时，情况有何不同?

进一步阅读 定理(6.3.5)的原始形式可参看 A. J. Hoffman and H. Wielandt, "The Variation of the Spectrum of a Normal Matrix," *Duke Math. J.* 20(1953), 37-39. 这个结果关于实对称情形的一个初等证明见[Wil], pp. 104-109. 377

6.4 其他包含区域

我们已经比较详细地讨论了 Geršgorin 圆盘. 它们是平面上一类特殊的容易计算的区域，这些区域保证包含给定矩阵的各特征值. 许多作者也许受 Geršgorin 理论的优美几何结构的启发，推广了这个理论的思想和方法，得到一些其他类型的包含区域. 我们讨论其中几个已经成熟的富有特色的结果.

第一个结果给出了特征值的包含区域，像 Geršgorin 区域那样，它是若干个圆盘之并，不过，圆盘的半径既依赖于去心行和也依赖于去心列和. Geršgorin 定理的分离行和与列和形式是作为这个结果的极限而得到的，这个结果归功于 Ostrowski，因此它可以看成安插在(6.1.2)与(6.1.4)之间的包含区域的一个闭联集.

6.4.1 定理(Ostrowski) 设 $A=[a_{ij}]\in M_n$，$\alpha\in[0,1]$是给定的数，又设 R_i' 和 C_i' 分别表示 A 的去心行和与去心列和：

$$R_i'=\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}| \tag{6.4.2}$$

$$C_i'=\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ji}| \tag{6.4.3}$$

则 A 的所有特征值位于 n 个圆盘的并

$$\bigcup_{i=1}^{n}\{z\in\mathbf{C}:|z-a_{ii}|\leqslant R_i'^{\alpha}C_i'^{1-\alpha}\} \tag{6.4.4}$$

中.

证明：我们可以假定 $0<\alpha<1$，当 $\alpha=0$ 和 $\alpha=1$ 的情形(相应于关于列和与行和的 Geršgorin 定理)可以通过取极限得到. 另外，可以假定所有 $R_i'>0$，因为可以在其 $R_i'=0$ 的任一行中添上一个小的非零元使 A 产生扰动；所得到的矩阵有包含区域(6.4.4)，它大于 A 的包含区域，并且在扰动趋于零的极限情形推出结论成立.

378 现在假定 $Ax=\lambda x$，且 $x=[x_i]\neq 0$. 于是对每个 $i=1, 2, \cdots, n$，有

$$|\lambda-a_{ii}||x_i|=\left|\sum_{\substack{j=1\\j\neq i}}^{n}a_{ij}x_j\right|\leqslant\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_j|=\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|^{\alpha}\{|a_{ij}|^{1-\alpha}|x_j|\}$$

$$\leqslant\left[\sum_{\substack{j=1\\j\neq i}}^{n}\{|a_{ij}|^{\alpha}\}^{1/\alpha}\right]^{\alpha}\left[\sum_{\substack{j=1\\j\neq i}}^{n}\{|a_{ij}|^{1-\alpha}|x_j|\}^{1/(1-\alpha)}\right]^{1-\alpha} \tag{6.4.5a}$$

$$=R_i'^{\alpha}\left[\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_j|^{1/(1-\alpha)}\right]^{1-\alpha},$$

因为 $R_i'>0$，它等价于

$$\frac{|\lambda-a_{ii}|}{R_i'^{\alpha}}|x_i|\leqslant\left[\sum_{j=1}^{n}|a_{ij}||x_j|^{1/(1-\alpha)}\right]^{1-\alpha},$$

因而

$$\left[\frac{|\lambda-a_{ii}|}{R_i'^{\alpha}}\right]^{1/(1-\alpha)}|x_i|^{1/(1-\alpha)}\leqslant\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_j|^{1/(1-\alpha)}. \tag{6.4.5b}$$

在(6.4.5a)中所使用的不等式是 Holder 不等式(附录 B)，其中，$p=1/\alpha$ 且 $q=p/(p-1)=1/(1-\alpha)$. 现在对 i 相加(6.4.5b)便得到

$$\sum_{i=1}^{n}\left[\frac{|\lambda-a_{ii}|}{R_i'^{\alpha}}\right]^{1/(1-\alpha)}|x_i|^{1/(1-\alpha)}\leqslant\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_j|^{1/(1-\alpha)}$$

$$=\sum_{j=1}^{n}C_j'|x_j|^{1/(1-\alpha)}. \tag{6.4.6}$$

如果对适合 $x_i\neq 0$ 的每个 i 有

$$\left[\frac{|\lambda-a_{ii}|}{R_i'^{\alpha}}\right]^{1/(1-\alpha)}>C_i',$$

则(6.4.6)不能成立. 因此，得出

$$\left[\frac{|\lambda-a_{ii}|}{R_i'^{\alpha}}\right]^{1/(1-\alpha)}\leqslant C_i'$$

对适合 $x_i\neq 0$ 的至少一个 i 值成立，因而

$$|\lambda-a_{ii}|\leqslant R_i'^{\alpha}C_i'^{1-\alpha}.$$

□

练习 考虑 $A=\begin{bmatrix}1 & 4\\1 & 6\end{bmatrix}$，试比较 Geršgorin 行和列特征值包含区域与 $\alpha=\frac{1}{2}$ 的 Ostrowsi 区

域. 对于 A 的谱半径，Ostrowski 定理给出什么估计，且如何同 Geršgorin 估计(6.1.5)进行比较?

379

练习　推论(6.1.6)的 Ostrowski 形式是什么?

下一个结果也是 Geršgorin 定理的推广，它属于 Brauer，不过现在一次要取两行；几何区域不再是圆盘，而是称作 Cassini 椭圆形的诸集合. 显然，其证明平行于 Geršgorin 定理的证明，其中不仅选择特征向量的一个最大模分量，而且还选择两个最大模分量.

6.4.7 定理(Brauer)　设 $A=[a_{ij}]\in M_n$. A 的所有特征值位于 $n(n-1)/2$ 个 Cassini 椭圆形的并

$$\bigcup_{\substack{i,j=1\\ i\neq j}}^{n}\{z\in\mathbf{C}:|z-a_{ii}||z-a_{jj}|\leqslant R_i'R_j'\} \tag{6.4.8}$$

中.

证明： 设 λ 是 A 的特征值，且假定 $Ax=\lambda x$，其中 $x=[x_i]\neq 0$. 则 x 的一个元素有最大绝对值，例如 x_p，使得对所有 $i=1,\cdots,n$ 有 $|x_p|\geqslant|x_i|$ 且 $x_p\neq 0$. 如果 x 的所有其他元都是零，则 $Ax=\lambda x$ 的假定说明 $a_{pp}=\lambda$. 因为 A 的所有对角元都包含在区域(6.4.8)中，所以，当它的相应特征向量只有一个非零元时，特征值 λ 在这个区域中.

现在假定至少存在特征向量 x 的两个非零元，设 x_q 是具有第二个最大绝对值的分量；即对所有 $i=1,\cdots,n$，$i\neq p$，且 $x_p\neq 0\neq x_q$，有 $|x_p|\geqslant|x_q|\geqslant|x_i|$.

于是 $Ax=\lambda x$ 表明

$$x_p(\lambda-a_{pp})=\sum_{\substack{j=1\\ j\neq p}}^{n}a_{pj}x_j$$

它蕴涵

$$|x_p||\lambda-a_{pp}|=\Big|\sum_{\substack{j=1\\ j\neq p}}^{n}a_{pj}x_j\Big|\leqslant\sum_{\substack{j=1\\ j\neq p}}^{n}|a_{pj}||x_j|\leqslant\sum_{\substack{j=1\\ j\neq p}}^{n}|a_{pj}||x_q|=R_p'|x_q|,$$

或

$$|\lambda-a_{pp}|\leqslant R_p'\frac{|x_q|}{|x_p|}. \tag{6.4.9}$$

但是也有

$$x_q(\lambda-a_{qq})=\sum_{\substack{j=1\\ j\neq q}}^{n}a_{qj}x_j$$

它蕴涵

380

$$|x_q||\lambda-a_{qq}|=\Big|\sum_{\substack{j=1\\ j\neq q}}^{n}a_{qj}x_j\Big|\leqslant\sum_{\substack{j=1\\ j\neq q}}^{n}|a_{qj}||x_j|\leqslant\sum_{\substack{j=1\\ j\neq q}}^{n}|a_{qj}||x_p|=R_q'|x_p|,$$

或

$$|\lambda-a_{qq}|\leqslant R_q'\frac{|x_p|}{|x_q|}. \tag{6.4.10}$$

取(6.4.9)和(6.4.10)的乘积使我们可消去 x 的两个分量的未知比，从而得到

$$|\lambda - a_{pp}||\lambda - a_{qq}| \leqslant R'_p \frac{|x_q|}{|x_p|} R'_q \frac{|x_p|}{|x_q|} = R'_p R'_q.$$

因此，特征值 λ 位于区域(6.4.8)中. □

练习 Brauer 定理的列和形式是什么？

关于特征值包含区域的任一个定理蕴涵(并且实际上也被蕴涵于)相关的可逆性定理. 现在利用包含结果得到不许 $z=0$ 属于该包含区域的条件.

6.4.11 推论 如果 $A=[a_{ij}]\in M_n$，则下列条件中的任何一个都是 A 为可逆矩阵的充分条件：

(a) 对某个 $\alpha\in[0, 1]$，$|a_{ii}| > R_i'^{\alpha} C_i'^{1-\alpha}$ 对所有 $i=1, \cdots, n$ 成立(Ostrowskii)；

(b) $|a_{ii}||a_{jj}| > R'_i R'_j$ 对所有 $i, j=1, \cdots, n$，$i\neq j$ 成立(Brauer).

练习 利用(6.4.1)和(6.4.7)证明(6.4.11).

Brauer 定理要求一次取两行的乘积. 在一次取三行或多行的想法启发下，人们注意到进一步推广 Brauer 定理的可能性，并且对每个 $m=1, \cdots, n$，考虑形如

$$\left\{z \in \mathbf{C}: \prod_{k=1}^{m} |z - a_{i_k i_k}| \leqslant \prod_{k=1}^{m} R'_{i_k}\right\}, \quad A = [a_{ij}] \in M_n \tag{6.4.12}$$

的集合之并. 对每个 m，有 $\binom{n}{m}$ 个这种形式的集合：$m=1$ 给出 n 个 Geršgorin 圆盘，而 $m=2$ 给出 Brauer 的 $n(n-1)/2$ 个椭圆形. 遗憾的是，对 $m\geqslant 3$，集合(6.4.12)并不一定是特征值包含区域，正如例子

[381]

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.4.13}$$

所说明的那样. 对于 $m=3$ 和 $m=4$，集合(6.4.12)在点 $z=2$ 完全失效.

练习 证明(6.4.13)中矩阵的特征值是 $\lambda=0, 1, 1$ 和 2. 对 $m=1$，$m=2$ 和 $m=3$，4 画出集合(6.4.12)的草图. 考虑

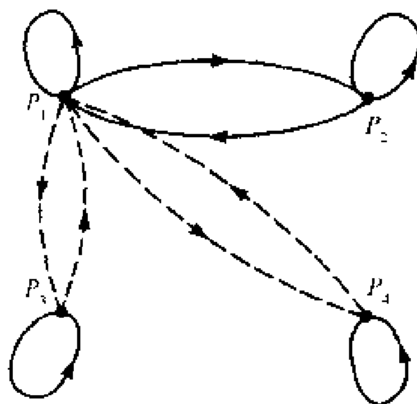
$$A = \begin{bmatrix} J & 0 \\ 0 & I_n \end{bmatrix} \in M_{n+2}, \tag{6.4.14}$$

其中，$J=\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\in M_2$，且 $I_n\in M_n$ 是 $n\times n$ 单位矩阵，证明，对所有 $m\geqslant 3$，也会出现上述同样的现象.

虽然这个例子排除了对 Brauer 定理作最明显的推广的可能性，但是它提醒我们，什么样的推广可能是错误的，应如何对待. 区域(6.4.12)的问题在于采用了过多项的乘积，其中一些项因为零去心行和而有可能为零. 当然，如果矩阵 A 是不可约的，这种情况不会出能现；这时所有 $R'_i>0$.

然而，即使 A 是不可约的，区域(6.4.12)还可能不是 A 的特征值包含区域；它仍然可能采用很多项的乘积. 考虑由

$$A_\varepsilon=\begin{bmatrix}1&1&\varepsilon&\varepsilon\\1&1&0&0\\\varepsilon&0&1&0\\\varepsilon&0&0&1\end{bmatrix},\quad 1>\varepsilon\geqslant 0,\tag{6.4.15}$$



382

给出的扰动及 A_ε 的有向图 $\Gamma(A_\varepsilon)$，其中，当 $\varepsilon=0$ 时，虚弧线消失. 如果 $\varepsilon\neq0$，则 $\Gamma(A_\varepsilon)$ 是强连通的，且 A_ε 是不可约的. 于是求得

$$R_1'=1+2\varepsilon,\quad R_2'=1,\quad R_3'=\varepsilon,\quad R_4'=\varepsilon,$$

且 A_ε 有特征值

$$\lambda_\varepsilon=1,\ 1,\ 1+(1+2\varepsilon^2)^{1/2}\ \text{和}\ 1-(1+2\varepsilon^2)^{1/2}.$$

练习 验证关于 A_ε 的上述计算.

因为三项或多项 R_i' 的任一乘积至少包含 ε 的一个因子，所以，当 ε 是小的正数时，无论对于 $m=3$ 还是 $m=4$，集合(6.4.12)都不可能是特征值的包含区域.

练习 考虑(6.4.14)的如同形式(6.4.15)的扰动，证明上述结论对所有 $m\geqslant3$ 成立.

(6.4.13)和(6.4.15)的哪个固有性质说明，$m=1$ 和 $m=2$ 在(6.4.12)中是可行的，而 $m=3$ 和 $m=4$ 是不可行的呢？Richard Brualdi 注意到，在每种情形下有向图都不包含长为 3 或 4 的回路，但是的确包含长为 1 和 2 的回路. 这原来是得到 Brauer 定理的正确推广的关键所在.

我们知道，有向图 Γ 是强连通的，当且仅当在 Γ 中有一条从每一结点到任何其他结点的有向道路(并且还有返回的有向道路).

我们称 Γ 是弱连通的，当且仅当从每个结点到某个其他结点有一条有向道路及返回的有向道路. 这等价于断言，Γ 中的每个结点属于某条非平凡回路；一条平凡回路(或圈)是长为 1 的有向道路，其起点和终点在同一个结点.

用矩阵来描述，我们知道，$\Gamma(A)$ 是强连通的，当且仅当 A 是不可约的. 称 A 是弱不可约的，当且仅当 $\Gamma(A)$ 是弱连通的. 弱不可约性看来不像不可约性有引人注意的特征描述[用像置换相似那样的变换表示(6.2.21b)]，但是，用 A 元素的零-非零结构来描述便会得知，A 是弱不可约的，当且仅当对每个 $i=1,\cdots,n$，A 的第 i 行至少有一个非零非对角元 a_{ij_i}，使得 A 有一个非零元序列 $a_{k_1k_2},a_{k_2k_3},\cdots,a_{k_{m-1}k_m}$，适合 $k_1=j_i$ 和 $k_m=i$. 这个麻烦的条件大约是 A 有性质 SC 的条件(6.2.7)的一半，为了计算上的需要，或许用类似于定理(6.2.23)的形式来叙述更为方便.

383

6.4.16 引理 如果 $A\in M_n$，则 A 是弱不可约的，当且仅当矩阵

(a) $B=[I+|A|]^{n-1}$ 或

(b) $B=[I+M(A)]^{n-1}$

中的任何一个有性质：对每个 $i=1,\cdots,n$，在第 i 行至少存在一个非零非对角元 $b_{ij}\ (j\neq i)$，

使得 b_n 也是非零的.

练习 证明引理(6.4.16). **提示:** 利用(6.2.19)中的思想

练习 假定 $A\in M_n$，设 $B\in M_n$ 定义为(6.4.16)中的(a)或(b). 证明，A 是弱不可约的当且仅当 $\Gamma(B)$ 有性质：每个结点属于长为 2 的回路. 关于不可约矩阵的相应性质是什么？哪个性质更弱？根据定义可知，一个回路是简单的，如果只有它的起点(这也是终点)在结点表中可以出现两次以上.

练习 如果 $A\in M_n$ 是弱不可约的，证明所有 $R_i'>0$ 和所有 $C_i'>0$.

集 S 上的一个预序是定义在 S 的所有点对之间的关系 R，使得对任一对元素 s，$t\in S$，有 sRt 或 tRs，或者两者都成立. 一个预序一定是自反的(sRs 对每个 $s\in S$ 成立)和传递的(如果 sRt 且 tRu，则 sRu). 一个预序可能不是对称的(只要 tRs，就有 sRt)，而且还可能有 sRt 且 tRs，而无 $s=t$. S 的子集 S_0 中的点 z 称为 S_0 的极大元，是指 sRz 对所有 $s\in S$ 成立.

练习 设 S 是复数域的任一非空集合. 用

$$zRw,\quad \text{当且仅当} |z|\leqslant|w|$$

定义的复数对 z，$w\in S$ 之间的关系是 $\mathbf{C}$ 上一个预序.

6.4.17 引理 设 S 是非空有限集，在 S 上定义了一个预序. 则 S 至少包含一个极大元.

证明: 把诸元素排成任意顺序 s_1，…，s_k，令 $s\equiv s_1$. 如果 s_2Rs，就留下 s，否则就令 $s=s_2$. 对余下的元素继续这样做. s 的最后值就是极大元. □

384

如果 Γ 是有向图，且 P_i 是 Γ 的一个结点，定义 $\Gamma_{外}(P_i)$ 是由不同于 P_i 的结点组成的集，且从 P_i 经长为 1 的某条有向道路可以到达其中每一个结点. 注意，如果 Γ 是弱连通的，则对每个结点 $P_i\in\Gamma$，$\Gamma_{外}(P_i)$ 是不空的.

让我们用 $C(A)$ 表示有向图 $\Gamma(A)$ 中的诸非平凡回路 γ 的集合. 一条非平凡回路是至少包含两个不同结点的回路；即它是(简单有向)回路，而不是圈. 对于矩阵(6.4.13)，$C(A)$ 只包括一条回路 $\gamma=P_1P_2$，P_2P_1. 而对于矩阵(6.4.15)，则有三条分离的非平凡回路，其长度都是 2.

6.4.18 定理(Brualdi) 如果 $A=[a_{ij}]\in M_n$ 是弱不可约的，则 A 的每个特征值包含在区域

$$\bigcup_{\gamma\in C(A)}\left\{z\in\mathbf{C}:\prod_{P_i\in\gamma}|z-a_{ii}|\leqslant\prod_{P_i\in\gamma}R_i'\right\}\tag{6.4.19}$$

中. 记号是说，如果 $\gamma=P_{i_1}P_{i_2}$，…，$P_{i_k}P_{i_{k+1}}$ 是非平凡回路，且 $P_{i_{k+1}}=P_{i_1}$，则在(6.4.19)中的每个乘积恰好包含 k 项，且下标 i 取 i_1，i_2，…，i_k 这 k 个值.

证明: 假定 λ 是 A 的特征值，且对 A 的某个主对角元有 $\lambda=a_{ii}$. 则 λ 显然在区域(6.4.19)中. 实际上，因为 A 是弱不可约的，所以所有 $R_i'>0$，因而在这种情形，λ 位于区域(6.4.9)内部. 如果 A 的每个特征值都等于 A 的某个主对角元，则 A 的所有特征值都在区域(6.4.9)的内部，于是定理得证.

关于其余情形的证明，假定 λ 是 A 的特征值，且对所有 $i=1$，…，n，$\lambda\neq a_{ii}$. 设 $Ax=\lambda x$ 对某个非零 $x=[x_i]\in\mathbf{C}^n$ 成立. 在 Γ 的诸结点上定义预序 R 为

$$P_iRP_j,\quad \text{当且仅当} |x_i|\leqslant|x_j|.\tag{6.4.20}$$

要证明，在 $\Gamma(A)$ 中存在回路 γ'，且具有以下三条性质：

$$\left.\begin{array}{l}\text{(a) } \gamma'=P_{i_1}P_{i_2},\ P_{i_2}P_{i_3},\ \cdots,\ P_{i_k}P_{i_{k+1}} \text{ 是非平凡(简单有向)回路,} \\ \quad\text{其中 } k\geqslant 2 \text{ 且 } P_{i_{k+1}}=P_{i_1}. \\ \text{(b) 对每个 } j=1,\ \cdots,\ k,\text{ 结点 } P_{i_{j+1}} \text{ 是 } \Gamma_{外}(P_{i_j}) \text{ 中的极大结点;} \\ \quad\text{即 } |x_{i_{j+1}}|\geqslant|x_m| \text{ 对所有使 } P_m\in\Gamma_{外}(P_{i_j}) \text{ 的 } m \text{ 成立.} \\ \text{(c) 所有 } x_{i_j}\neq 0,\ j=1,\ \cdots,\ k.\end{array}\right\} \tag{6.4.21}$$

[385]

如果 γ' 是适合条件(6.4.21)的回路，则 $Ax=\lambda x$ 推知，对任意 $j=1,\ \cdots,\ k$，有

$$(\lambda-a_{i_ji_j})x_{i_j}=\sum_{\substack{m=1\\m\neq i_j}}^{n}a_{i_jm}x_m=\sum_{P_m\in\Gamma_{外}(P_{i_j})}a_{i_jm}x_m,$$

因而

$$|\lambda-a_{i_ji_j}||x_{i_j}|=\left|\sum_{P_m\in\Gamma_{外}(P_{i_j})}a_{i_jm}x_m\right|\leqslant\sum_{P_m\in\Gamma_{外}(P_{i_j})}|a_{i_jm}||x_m| \tag{6.4.22}$$

$$\leqslant\sum_{P_m\in\Gamma_{外}(P_{i_j})}|a_{i_jm}||x_{i_{j+1}}|$$

$$=R'_{i_j}|x_{i_{j+1}}|. \tag{6.4.22a}$$

如果现在在 γ' 的所有结点上取不等式(6.4.22)的乘积，便得到

$$\prod_{j=1}^{k}|\lambda-a_{i_ji_j}||x_{i_j}|\leqslant\prod_{j=1}^{k}R'_{i_j}|x_{i_{j+1}}|. \tag{6.4.23}$$

但是

$$\prod_{j=1}^{k}|\lambda-a_{i_ji_j}|=\prod_{P_i\in\gamma'}|\lambda-a_{ii}| \quad 和 \quad \prod_{j=1}^{k}R'_{i_j}=\prod_{P_i\in\gamma'}R'_i,$$

又因为 $P_{i_{k+1}}=P_{i_1}$，还有 $x_{i_{k+1}}=x_{i_1}$. 因此，

$$\prod_{j=1}^{k}|x_{i_j}|=\prod_{j=1}^{k}|x_{i_{j+1}}|\neq 0. \tag{6.4.24}$$

于是，用(6.4.24)除(6.4.23)便得到

$$\prod_{P_i\in\gamma'}|\lambda-a_{ii}|\leqslant\prod_{P_i\in\gamma'}R'_i. \tag{6.4.25}$$

因为 γ' 是 $\Gamma(A)$ 中的非平凡回路，所以特征值 λ 一定位于区域(6.4.19)中.

现在应该证明，必存在适合条件(6.4.21)的回路 γ'. 设 i 是使 $x_i\neq 0$ 的任一下标；于是从恒等式

$$(\lambda-a_{ii})x_i=\sum_{\substack{j=1\\j\neq i}}^{n}a_{ij}x_j=\sum_{P_j\in\Gamma_{外}(P_i)}a_{ij}x_j,$$

以及 $x_i\neq 0$ 和 $\lambda-a_{ii}\neq 0$ 的事实得知，左边非零，因而在 $\Gamma_{外}(P_i)$ 的诸结点中[这些 P_j 使 $a_{ij}\neq 0$ 且 $j\neq i$；因为 $\Gamma(A)$ 是弱连通的，所以 $\Gamma_{外}(P_i)$ 非空]一定至少有一个结点，使得相应的特征向量分量 x_j 非零. 设 $P_{i_1}\equiv P_i$，又设 P_{i_2} 是 $\Gamma_{外}(P_{i_1})$ 中的诸结点中的极大结点，即 $|x_{i_2}|\geqslant|x_m|$ 对所有使 $P_m\in\Gamma_{外}(P_{i_1})$ 的 m 成立. 我们可保证 $x_{i_2}\neq 0$.

[386]

假定上述构造法已经得到长为 $j-1$ 的有向道路 $P_{i_1}P_{i_2}$，$P_{i_2}P_{i_3}$，…，$P_{i_{j-1}}P_{i_j}$，且适合(6.4.21)的条件(b)和(c)；刚才已对 $j=2$ 这样做了．于是

$$(\lambda-a_{i_ji_j})x_{i_j}=\sum_{P_m\in\Gamma_{外}(P_{i_j})}a_{i_jm}x_m,$$

且左边非零，因而在 $\Gamma_{外}(P_{i_j})$中一定至少存在一个结点[因 $\Gamma(A)$是弱连通的，所以 $\Gamma_{外}(P_{i_j})$非空]使相应的特征向量分量非零．因此，如果选取 $P_{i_{j+1}}$ 为 $\Gamma_{外}(P_{i_j})$的极大结点，保证有 $x_{i_{j+1}}\neq 0$.

因为在 $\Gamma(A)$中只有有限多个结点，这个构造法依次取 $j=2$，3，…，最后会得到第一个极大结点 $P_{i_q}\in\Gamma_{外}(P_{i_{q-1}})$．它是作为前面某一步的结点 P_{i_p} 得来的$(2\leqslant p+1<q)$．因此 $\gamma'=P_{i_p}P_{i_{p+1}}$，$P_{i_{p+1}}P_{i_{p+2}}$，…，$P_{i_{q-1}}P_{i_q}$ 是$\Gamma(A)$中的回路，且适合(6.4.21)中的所有三个条件． □

当 A 实际上是不可约的时候，Brualdi 定理有较强的形式，它是推广 Brauer(6.4.7)的定理(6.2.26)的形式．

6.4.26 定理(Brualdi) 设 $A=[a_{ij}]\in M_n$ 是不可约的．则对每个非平凡回路 $\gamma\in C(A)$，只有当各个集合

$$\left\{z\in\mathbf{C}:\prod_{P_i\in\gamma}|z-a_{ii}|\leqslant\prod_{P_i\in\gamma}R_i'\right\}\tag{6.4.27}$$

的边界都经过 λ 时，区域(6.4.19)的边界点 λ 才可以是 A 的特征值．

证明： 因为所有 $R_i'>0$，如果 $\lambda=a_{ii}$ 对任一 $i=1$，2，…，n 成立，则 λ 不可能在区域(6.4.27)的边界上．因此，可以假定，对所有 $i=1$，…，n 有 $\lambda\neq a_{ii}$，并且可以继续采用 Brualdi 定理(6.4.18)中的证法以及相同的记号．但是需另外假定 λ 是在区域(6.4.19)的边界上的 A 的特征值．正如在引理(6.2.3)的证明中所证明的，对所的有非平凡回路 $\gamma\in C(A)$，λ 必定满足不等式

$$\prod_{P_i\in\gamma}|\lambda-a_{ii}|\geqslant\prod_{P_i\in\gamma}R_i'$$

387 [其中等式对至少一个 $\gamma\in C(A)$成立]．把这个不等式与(6.4.25)作一比较，得知

$$\prod_{P_i\in\gamma'}|\lambda-a_{ii}|=\prod_{P_i\in\gamma'}R_i',\tag{6.4.28}$$

其中 γ'是在(6.4.18)的证明中所构造的特殊回路．因此，(6.4.23)中的不等式一定是等式，从而对所有 $j=1$，2，…，k，(6.4.22)中的两个不等式一定是等式．特别是，不等式(6.4.22a)一定是等式，因而对每个 $P_{i_j}\in\gamma'$和对所有使 $P_m\in\Gamma_{外}(P_{i_j})$的 m，$|x_m|=|x_{i_{j+1}}|=c_{i_{j+1}}$ 一常数．注意，这个结论对适合条件(6.4.21)的任一回路都成立．

现在定义集合

$$K\equiv\{P_i\in\Gamma(A):|x_m|=c_i=\text{常数对所有使 }P_m\in\Gamma_{外}(P_i)\text{ 的 }m\text{ 成立}\}.$$

因为 γ'的所有结点都在 K 中，所以 K 不空．我们想要证明 $\Gamma(A)$的所有结点都在 K 中．

假定 $\Gamma(A)$的结点 P_q 不在 K 中．因为 $\Gamma(A)$是强连通的，所以在 $\Gamma(A)$中，从 K 的每个结点到这个外结点 P_q 至少存在一条有向道路．如果从所有这些有向道路中选取长度最小的道路，则它的第一条弧一定是从 K 中的一个结点到不在 K 中的结点 P_r 的弧．如果在 $\Gamma(A)$的结点上

采用定理(6.4.18)的证明中用过的相同预序，则可以使用在定理(6.4.18)证明中用过的相同构造法：从结点 $P_i \equiv P_{j_1}$ 开始，选取极大结点 $P_{j_2} \in \Gamma_{外}(P_{j_1})$，选取极大结点 $P_{j_3} \in \Gamma_{外}(P_{j_2})$，如此等等. 因为 $\Gamma(A)$ 是弱(甚至强)连通的，所以，在每一步 $\Gamma_{外}(P_{j_i})$ 非空，又由于如前所述的相同理由，极大结点适合(6.4.21)的条件(c).

如果在这个构造过程中的某一步，在选取极大结点当中，可以选择在 K 中的或不在 K 中的结点，那么就总选取不在 K 中的那个结点. 如果在任一步中，可供选取的所有极大结点都在 K 中，那么就在它们中任选一个，然后沿着(一定在 K 中的)长度最小的有向道路到不在 K 中的第一个结点上；并且如前继续选择极大结点. 根据 K 的定义，K 中任一条有向道路将具有性质：每个结点是它的前一结点在 $\Gamma_{外}$ 中的极大结点[(6.4.21)的条件(b)]. 因为 K 的补集只有有限多个结点，这个构造法终究应该得到 K 的补集中的第一个极大结点，它是作为前面某一步的一个结点得来的. 对于这个构造法中的这个结点，在它第一次出现与第二次出现之间的有向道路将是一条非平凡的有向回路，这条有向回路可能不是简单的，这是因为当这个构造法选出 K 中一结点时，我们就选一条由该点到 K 外一结点的有向道路. 在位于 K 中的那部分道路中，可能有有限多条回路，但是可以删去它们而留下一条简单回路 γ''，γ'' 适合条件(6.4.21)且至少包含不在 K 中的一个结点.

388

因为回路 γ'' 适合条件(6.4.21)，所以可以用它来代替定理(6.4.18)证明中的回路 γ'. 根据本证明的第一段中的论证，得知，对所有 $P_{j_i} \in \gamma''$，$|x_m| = c_{j_i} =$ 常数对所有 $P_m \in \Gamma(P_{j_i})$ 成立. 因此 γ'' 中的每个结点在 K 中，这与 γ'' 至少包含不在 K 中的一个结点相矛盾. 这就证明了 $\Gamma(A)$ 不可能有不在 K 中的任何结点.

如果 γ 是 $\Gamma(A)$ 中任一条非平凡(简单有向)回路，因为它的所有结点都在 K 中，所以它自然适合条件(6.4.21). 因此可以用它代替定理(6.4.18)的证明中的 γ'，从而也可以用它代替(6.4.28)中的 γ'. 这就是所要证的结论：每个集(6.4.27)的边界经过 λ. □

6.4.29 推论 如果 $A \in M_n$，则下列条件中的任何一个是 A 为可逆矩阵的充分条件：

(a) A 是弱不可约矩阵，且

$$\prod_{P_i \in \gamma} |a_{ii}| > \prod_{P_i \in \gamma} R_i'$$

对每条非平凡回路 $\gamma \in C(A)$ 成立；

(b) A 是不可约矩阵，且

$$\prod_{P_i \in \gamma} |a_{ii}| \geqslant \prod_{P_i \in \gamma} R_i'$$

对每条非平凡回路 $\gamma \in C(A)$ 成立，其中严格不等式对至少一条回路成立.

习题

1. 证明，如果矩阵 $A=[a_{ij}]$ 适合关于 Brauer 可逆性的条件(6.4.11b)，则除了 $i=1,\cdots,n$ 的至多一个值以外，都有 $|a_{ii}|>R_i'$. 因此，Brauer 条件只是比关于严格对角占优性的 Levy-Desplanques 条件(6.1.10a)稍微弱一些. 这与(6.1.11)是什么关系？

2. 证明，根据(6.4.11)中的两个条件，$A=\begin{bmatrix}2 & 3\\ 1 & 3\end{bmatrix}$ 都是可逆的，不过不是 Levy-

Desplanques条件(6.1.10a)也不是(6.1.11)确保其可逆性.(6.1.11)的列形式是什么?

3. 证明，对于 $n\geqslant 2$，每个不可约矩阵 $A\in M_n$ 是弱不可约的. 试给出一个例子，它是弱不可约矩阵而不是不可约矩阵.

389

4. 给出推论(6.4.29)的证明细节. **提示**：采用(6.1.10)和(6.2.6)中的相同证法.

5. 证明，$A\in M_n$ 是弱不可约的，当且仅当 A 不置换相似于其一个对角子块是 1×1 的分块三角(0.9.4)矩阵.

进一步阅读 关于包含区域的详细说明以及许多原始文献可参看 R. Brualdi, "Matrices, Eigenvalues, and Directed Graphs," *Lin. Multilin. Alg.* 11(1982), 143-165.

390

第7章 正定矩阵

7.0 导引

一类具有特殊正性的 Hermite 矩阵常常出现在许多应用中．具有这种正性的 Hermite 矩阵（特别是实对称矩阵）可以看作正数概念到矩阵的推广．这样考虑常常可以深入理解正定矩阵的一些性质和应用．下面给出这方面的一些例子，其中就要出现这些特殊的 Hermite 矩阵．

Hessian 矩阵，极小化和凸性

设 $f(x)$ 是某区域 $D\subset\mathbf{R}^n$ 上的光滑实值函数．如果 $y=[y_i]$ 是 D 的一个内点，则 Taylor 定理说明，

$$f(x)=f(y)+\sum_{i=1}^{n}(x_i-y_i)\left.\frac{\partial f}{\partial x_i}\right|_y+\sum_{i,j=1}^{n}(x_i-y_i)(x_j-y_j)\left.\frac{\partial^2 f}{\partial x_i\,\partial x_j}\right|_y+\cdots$$

对 y 附近的点 $x\in D$ 成立，如果 y 是 f 的临界点，则所有一阶偏导数在 y 点为零，因而，关于 f 在 y 附近的性态，有表示式

$$\begin{aligned}f(x)-f(y)&=\sum_{i,j=1}^{n}(x_i-y_i)(x_j-y_j)\left.\frac{\partial^2 f}{\partial x_i\,\partial x_j}\right|_y+\cdots\\&=(x-y)^T H(f;y)(x-y)+\cdots,\end{aligned}$$

391

$m\times n$ 矩阵

$$H(f;y)=\left[\left.\frac{\partial^2 f}{\partial x_i\,\partial x_j}\right|_y\right]$$

称为 f 在 y 点的 Hessian 矩阵；因为 f 的混合偏导数相等，所以它是对称矩阵．如果二次型

$$z^T H(f;y)z,\quad z\neq 0, z\in\mathbf{R}^n \tag{7.0.1}$$

总是正的，则 y 是 f 的相对极小点，如果这个二次型总是负的，则 y 是 f 的相对极大点．当然，如果这个二次型对所有非零 $z\in\mathbf{R}^n$ 可能没有确定的符号，在这种情况下，临界点 y 的性质就不确定．在 $n=1$ 的情形，验证相对极小点或极大点的这些准则不过是通常的二阶导数检验法．对于 $n=1$，第三种可能性只在拐点出现；当 $n>1$ 时，情况可能要复杂得多．

如果二次型(7.0.1)在 D 的所有点(不只是在 f 的临界点)非负，则 f 是 D 中的凸函数．这是 $n-1$ 的熟悉情形的直接推广．

方差-协方差矩阵

设 X_1，X_2，…，X_n 是在某个具有期望函数 E 的概率空间上的具有有限二阶矩的实或复随机变量，且假定 $\mu_i=E(X_i)$ 是相应的平均值．随机向量 $X=(X_1,\cdots,X_n)^T$ 的协方差矩阵是矩阵 $A=[a_{ij}]$，其中

$$a_{ij}=E[(X_i-\mu_i)(X_j-\mu_j)],\quad i,j=1,\cdots,n.$$

显然，A 是 Hermite 矩阵，并且容易算出，如果 $z=[z_i]\in\mathbf{C}^n$，则

$$z^*Az=E\Big[\sum_{i,j=1}^n \bar{z}_i(X_i-\bar{\mu}_i)z_j(X_j-\mu_j)\Big]=E\Big|\sum_{i=1}^n z_i(X_i-\mu_i)\Big|^2\geqslant 0.$$

在这个结论中所涉及到的期望函数的仅有性质是它的线性、齐性和非负性三个性质；即只要 Y 是非负随机变量，就有 $E[Y]\geqslant 0$.

不借助概率术语也可以作出同样的结论. 如果在直线上有一族复值函数 $f_1, f_2, \cdots, f_n$，

392

如果 g 是实值函数，又如果所有积分

$$a_{ij}=\int_{-\infty}^{\infty}\bar{f}_i(x)f_j(x)g(x)\mathrm{d}x,\quad i,j=1,\cdots,n$$

都有定义且收敛，则矩阵 $A=[a_{ij}]$ 显然是 Hermite 矩阵. 容易算出，

$$z^*Az=\sum_{i,j=1}^n\int_{-\infty}^{\infty}\bar{z}_i\bar{f}_i(x)z_jf_j(x)g(x)\mathrm{d}x=\int_{-\infty}^{\infty}\Big|\sum_{i=1}^n z_if_i(x)\Big|^2g(x)\mathrm{d}x.$$

因此，如果 $g(x)$ 是非负函数，则这个二次型就是非负的.

非负函数的代数矩

设 $f(x)$ 是单位区间 $[0, 1]$ 上的绝对可积实值函数，并且考虑数

$$a_k\equiv\int_0^1 x^kf(x)\mathrm{d}x.\tag{7.0.2}$$

序列 $a_0, a_1, a_2, \cdots$ 称为 Hausdorff 矩序列，且它自然与实二次型

$$\sum_{j,k=0}^n a_{j+k}z_jz_k=\sum_{j,k=0}^n\int_0^1 x^{j+k}z_jz_kf(x)\mathrm{d}x=\int_0^1\Big(\sum_{k=0}^n z_kx^k\Big)^2f(x)\mathrm{d}x\tag{7.0.3}$$

有联系. 如果令 $A=[a_{i+j}]$，则 A 就是实对称矩阵，又如果对所有 $x\in[0, 1]$ 有 $f(x)\geqslant 0$，则对所有 $Z\in\mathbf{R}^{n+1}$，将有 $z^TAz\geqslant 0$. 这对每个 $n=1, 2, \cdots$ 都成立，不论其二次型是否非负，具有 A 的结构的矩阵(即元素 a_{ij} 只是 $i+j$ 的函数)称为 Hankel 矩阵. 见(0.9.8)节.

非负函数的三角矩

设(θ)是$[0, 2\pi]$上的绝对可积实值函数，并且考虑数

$$a_k\equiv\int_0^{2\pi}e^{ik\theta}f(\theta)\mathrm{d}\theta,\quad k=\pm 1,\pm 2,\cdots.\tag{7.0.4}$$

393

序列 $a_0, a_1, a_{-1}, a_2, a_{-2}, \cdots$ 称为 Toeplitz 矩序列，并且它自然与二次型

$$\sum_{j,k=0}^n a_{j-k}\bar{z}_jz_k=\sum_{j,k=0}^n\int_0^{2\pi}e^{i(j-k)\theta}\bar{z}_jz_kf(\theta)\mathrm{d}\theta=\int_0^{2\pi}\Big|\sum_{k=0}^n z_ke^{-ik\theta}\Big|^2f(\theta)\mathrm{d}\theta\tag{7.0.5}$$

有联系，如果令 $A=[a_{i-j}]$，则 A 就是 Hermite 矩阵，又如果对所有 $\theta\in[0, 2\pi]$ 有 $f(\theta)\geqslant 0$，则对所有 $z\in\mathbf{C}^{n+1}$ 将有 $z^*Az\geqslant 0$. 这对每个 $n=1, 2, \cdots$ 都成立. 不论其二次型是否非负，实有 A 的结构的矩阵(即元素 a_{ij} 只是 $i-j$ 的函数)称为 Teoplitz 矩阵. 见(0.9.7)节. 事实上，对公式(7.0.4)稍加修改(其中，非负测度 d_μ 代替 $f(\theta)\mathrm{d}\theta$)，则二次型(7.0.5)是非负的，当且仅当 a_k 由修改后的公式得出(Bochner 定理).

关于微分方程的数值解的离散化和差分法

假定我们有形如

$$-y''(x)+\sigma(x)y(x)=f(x),\quad 0\leqslant x\leqslant 1,$$

$$y(0) = \alpha,$$
$$y(1) = \beta$$

的两点边值问题，其中，α 和 β 是给定的实常数，$f(x)$ 和 $\sigma(x)$ 是给定的实值函数. 如果我们把这个问题离散化，且只求 $y(kh) \equiv y_k$ 的值，$k=0, 1, \cdots, n+1$，又如果利用均差逼近导数项

$$y''(x) \cong \frac{y((k+1)h) - 2y(kh) + y((k-1)h)}{h^2} = \frac{y_{k+1} - 2y_k + y_{k-1}}{h^2},$$

那么，就得到线性方程组

$$\frac{-y_{k-1} + 2y_k - y_{k+1}}{h^2} + \sigma_k y_k = f_k, \quad k = 1,2,\cdots,n,$$

$$y_0 = \alpha,$$
$$y_{n+1} = \beta.$$

这里，对于正整数 n，可以取 $h=1/(n+1)$，$y_k=y(kh)$，$\sigma_k=\sigma(kh)$ 及 $f_k=f(kh)$. 可以把边值问题编入第一个($k=1$)方程和最后一个($k=n$)方程而给出方程组

394

$$(2 + h^2\sigma_1)y_1 - y_2 = h^2 f_1 + \alpha,$$
$$-y_{k-1} + (2 + h^2\sigma_k)y_k - y_{k+1} = h^2 f_k, \quad k = 2,3,\cdots,n-1,$$
$$-y_{n-1} + (2 + h^2\sigma_n)y_n = h^2 f_n + \beta,$$

还可以更紧凑地把它写成 $Ay=w$，其中 $y=[y_k]\in\mathbf{R}^n$，$w=[h^2f_1+\alpha, h^2f_2, \cdots, h^2f_{n-1}, h^2f_n+\beta]^T\in\mathbf{R}^n$，且 $A\in M_n$ 是三对角矩阵

$$A = \begin{bmatrix} 2+h^2\sigma_1 & -1 & & & 0 \\ -1 & 2+h^2\sigma_2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2+h^2\sigma_{n-1} & -1 \\ 0 & & & -1 & 2+h^2\sigma_n \end{bmatrix}. \qquad (7.0.6)$$

应该指出，不论 $\sigma(x)$ 为何值，A 都是实对称三对角矩阵，但是，如果希望 $Ay=w$ 对右边任意给定的值都是可解的，那么必须对 $\sigma(x)$ 作某些限制以保证 A 是非奇异矩阵.

容易算出相应于 A 的实二次型：

$$x^TAx = \left[x_1^2 + \sum_{i=1}^{n-1}(x_i - x_{i+1})^2 + x_n^2\right] + h^2\sum_{i=1}^{n}\sigma_i x_i^2.$$

等式右边中的括号内各项之和是非负的，且只有当 x 的各分量都相等，又都等于零时它才可能为零. 如果 $\sigma(x)\geqslant 0$，则后一个和式是非负的，且

$$x^TAx \geqslant \left[x_1^2 + \sum_{i=1}^{n-1}(x_i - x_{i+1})^2 + x_n^2\right] \geqslant 0. \qquad (7.0.7)$$

如果 A 是奇异矩阵，则存在某个非零向量 $\hat{x}\in\mathbf{R}^n$ 使得 $A\hat{x}=0$，因而 $\hat{x}^TA\hat{x}=0$. 另一方面，(7.0.7)中括号内各项之和必定为零，由此推出 $\hat{x}=0$. 因此，如果 $\sigma(x)\geqslant 0$，则矩阵 A 非奇异，且离散的边值问题对任意边界条件 α 和 β 都是可解的.

这是在研究常微分方程或偏微分方程的数值解时的典型情形. 为了计算上的稳定性，最好

是设计一种方法能把微分方程问题离散化，使得在所得到的线性方程组 $Ay=w$ 中 A 是正定矩阵．而当微分方程是椭圆型时，通常可以做到这一点．

395　在上述这些例子中所列举的矩阵具有特殊的正性，这正是本章要研究的对象．这些矩阵出现在许多应用中：调和分析中，复分析中，力学体系的振动理论中，以及矩阵理论的其他领域(例如奇异值分解和线性最小二乘方问题的解)中．

习题

1．如果序列 a_k 是由非负函数 f 通过公式(7.0.2)产生的，证明

$$\sum_{i,j=1}^{n} a_{i+j+1} z_i z_j \quad 和 \quad \sum_{i,j=1}^{n} \{a_{i+j}-a_{i+j+1}\} z_i z_j, \quad z=[z_i]\in \mathbf{R}^n$$

都是非负的．

2．用示意图说明 Hankel 矩阵中哪些对角线取常值．对 Toeplitz 矩阵作同样的说明．

3．证明，(7.0.6)中的矩阵 A 总是不可约的，又如果 $\sigma(x)\geqslant 0$，则它是不可约对角占优矩阵．试用推论(6.2.27)证明，A 是非奇异矩阵，且 A 的所有特征值是正数．

进一步阅读　关于实正定矩阵的一个简短的综述可参看 C. R. Johnson, "Positive Definite Matrices," *Amer. Math. Monthly* 77(1970), 259-264. 关于一般正定矩阵的其他综述以及有关这方面的大批参考资料可参看 O. Taussky, "Positive Definite Matrices," pp. 309-319 of *Inequalities*, ed. O. Shisha, Academic Press, New York, 1967；以及 O. Taussky, "Positive Definite Matrices and Their Role in the Study of the Characteristic Roots of General Matrices," *Advan. Math.* 2(1968), 175-186.

7.1　定义和性质

设 A 是 $n\times n$ Hermite 矩阵，如果对所有非零 $x\in \mathbf{C}^n$ 有

$$x^*Ax>0, \tag{7.1.1}$$

则称 A 为正定矩阵．如果(7.1.1)中所要求的严格不等式减弱成 $x^*Ax\geqslant 0$，则称 A 为半正定矩阵．在这些定义不等式中隐含着这样的事实：如果 A 是 Hermite 矩阵，则(7.1.1)左边总是实数．当然，如果 A 是正定矩阵，则它也是半正定矩阵．

396　**练习**　当 $n=1$ 时，正定矩阵和半正定矩阵意味着什么？

练习　证明，如果 $A\in M_n$，且对所有 $x\in \mathbf{C}^n$，x^*Ax 是实数，则 A 是 Hermite 矩阵．因此，A 是 Hermite 矩阵的假定在正定性定义中不是必不可少的．但是，一般习惯这样定义．**提示**：把 A 写成 $A=B+iC$，其中 B 和 C 是 Hermite 矩阵．

练习　说明，如果 $A\in M_n$ 是实矩阵，且对所有非零 $x\in \mathbf{R}^n$，x^TAx 是正数，则 A 不一定是对称矩阵．因而它不一定是正定矩阵．**提示**：考察实斜对称矩阵 A，然后计算 $(x^TAx)^T$．在这种情形 x^TAx 是什么？对于非实 x，x^*Ax 又如何呢？

练习　证明 $\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$ 是半正定的，但不是正定的．

练习　证明，如果 $A=[a_{ij}]\in M_n$ 是正定矩阵，则 $\bar{A}=[\bar{a}_{ij}]$，A^T，A^* 和 A^{-1} 也是正定矩阵．**提示**：如果 $Ay=x$，则 $x^*A^{-1}x=y^*A^*y$．

类似地，可以对 A 定义负定概念和半负定概念，这只要把正定和半正定的定义中的不等式颠倒一下即可，或等价地，分别把 $-A$ 定义为正定矩阵或半正定矩阵．因此，关于负定矩阵的任何一个命题都对应正定矩阵的一个相应命题．如果一个 Hermite 矩阵不属于上面提到的各类中的任何一个[即，如果(7.1.1)左边既可取正值也可取负值]，则称它为**不定矩阵**．

关于正定矩阵，可以作出几个直接的结论，并且对于半正定矩阵，也有类似的结论．

7.1.2 论断 正定矩阵的任一主子矩阵也是正定矩阵．

证明：设 S 是$\{1, 2, \cdots, n\}$的真子集，且用 $A(S)$表示从正定矩阵 $A\in M_n$ 中划去其行号和列号分别为 S 的补集的若干行和若干列后得到的矩阵．于是，$A(S)$是 A 的主子矩阵，且所有主子矩阵以这种形式出现；我们知道数 $\det A(S)$是 A 的主子式．设 $x\in \mathbf{C}^n$ 是这样一个非零向量，其中以 S 为下标的分量可取任意元，而其余分量为零元．设 $x(S)$表示从 x 中划去以 S 的补集为下标的(零)分量后得到的向量，因而得到

$$x(S)^* A(S)x(S) = x^* Ax > 0.$$

因为 $x(S)\neq 0$ 是任意的，这表明 $A(S)$是正定矩阵． □ 397

练习 证明正定矩阵的诸对角元是正实数．

7.1.3 论断 任意两个同阶正定矩阵的和是正定矩阵．更一般地，一些半正定矩阵的非负线性组合是半正定矩阵．

证明：设 A 和 B 是半正定矩阵，又设 a、$b\geqslant 0$，因而对任意 $x\in \mathbf{C}^n$ 有 $x^*(aA+bB)x=a(x^*Ax)+b(x^*Bx)\geqslant 0$. 多于两个矩阵的情形可按同样的方式处理．如果诸系数是正的，A 和 B 是正定矩阵，又如果 x 是非零向量，则和中的每一项都是正的，所以诸正定矩阵的正线性组合是正定矩阵． □

因此，正定矩阵的集合在所有矩阵组成的向量空间中是一个正锥．

7.1.4 论断 正定矩阵的每个特征值都是正实数．

证明：设 A 是正定矩阵，设 $\lambda\in\sigma(A)$，且 x 是 A 的相应于 λ 的特征向量．经计算，$x^*Ax=x^*\lambda x=\lambda x^*x$. 因此 $\lambda=(x^*Ax)/x^*x$ 是正数，因为它是两个正数之比． □

7.1.5 推论 正定矩阵的迹，行列式和所有主子式都是正数．

证明：迹与行列式恰好是诸特征值的和与积，余下的结论可由(7.1.2)推出． □

练习 证明半正定矩阵的诸特征值，迹，行列式和各主子式都是非负的．

练习 证明，$n\times n$ 负定矩阵的诸特征值和迹是负数，但是，对于奇数 n，其行列式是负数，对于偶数 n，其行列式是正数．

练习 证明，如果 $A=[a_{ij}]\in M_2$ 是正定矩阵，则 $a_{11}a_{22}>|a_{12}|^2$. **提示**：利用 $\det A>0$. 证明，如果 $A\in M_n$ 是正定矩阵，则对所有 i，$j=1, 2, \cdots, n$，$i\neq j$，

$$a_{ii}a_{jj} > |a_{ij}|^2.$$

证明，如果只假定 A 是半正定矩阵，则在上述不等式“$>$”必须换成“$\geqslant$”. 398

7.1.6 论断 设 $A\in M_n$ 是正定的．如果 $C\in M_{n,m}$，则 C^*AC 是半正定矩阵．另外，$\text{rank}(C^*AC)=\text{rank}(C)$，因而，$C^*AC$ 是正定矩阵当且仅当 C 有秩 m．

证明：首先指出，C^*AC 是 Herinite 矩阵. 对任意 $x\in\mathbf{C}^m$，有 $x^*C^*ACx=y^*Ay\geqslant 0$，其中 $y\equiv Cx$，而不等式是从 A 的正定性推出来的. 于是，C^*AC 是半正定矩阵. 另外，A 是正定矩阵，所以得知，$x^*C^*ACx>0$ 当且仅当 $Cx\neq 0$，只要能证明，$C^*ACx=0$ 当且仅当 $Cx=0$，则关于秩的论断(因而关于正定性的论断)也将得到证明，因为这将表明 C^*AC 和 C 有相同的零空间(因而它们也有相同的秩). 如果 $Cx=0$，则显然有 $C^*ACx=0$，反之，如果 $C^*ACx=0$，则 $x^*C^*ACx=0$，因而(如前，利用 A 的正定性)得出 $Cx=0$. □

练习 如果 $A\in M_n$ 是半正定矩阵而不是正定矩阵，又如果 $C\in M_n$，证明 C^*AC 总是半正定矩阵且不是正定矩阵. 如果 $C\in M_{n,m}$，且 $n\neq m$，用例子说明，即使 $A\in M_n$ 是奇异矩阵，C^*AC 也可能是正定矩阵.

练习 证明由正定(半正定)矩阵组成的锥在 * 相合下不变. 见(4.5.4).

练习 设 $A\in M_n$ 是 Hermite 矩阵. 证明，A 是正定(半正定)矩阵，当且仅当存在非奇异矩阵 $C\in M_n$ 使得 C^*AC 是正定(半正定)矩阵.

如果舍去 A 是 Hermite 的条件，并且定义的二次型(7.1.1)中只采用实变量，这会出现什么情况呢？如果 A 是具有实元素的矩阵，且 $x\in\mathbf{R}^n$，则 x^TAx 是实数，但我们仍然可能要问，对所有 $x\neq 0$，哪些矩阵有 $x^TAx>0$(即使 A 不是对称矩阵). 如果 A 是具有复元素的矩阵，或者如果允许 $x\in\mathbf{C}^n$，则可以用

$$\operatorname{Re}(x^*Ax)>0,\quad 其中所有非零\ x\in\mathbf{C}^n \tag{7.1.1'}$$

来代替(7.1.1). 定义 A 的 Hermite 部分为

$$H(A)\equiv\frac{1}{2}(A+A^*). \tag{7.1.7}$$

当 $n=1$ 时，这正好是复数 A 的实部.

练习 证明，(7.1.1′)成立，当且仅当 $H(A)$ 是正定矩阵.

练习 证明，对任意 $A\in M_n$，$A=H(A)+S(A)$，其中 $S(A)=\frac{1}{2}(A-A^*)$是 A 的斜 Hermite 部分.

399

习题

1. 设 $A\in M_n$ 是半正定矩阵且 $x\in\mathbf{C}^n$. 证明，$x^*Ax=0$ 的必要充分条件是 $Ax=0$. 由此推出，一个半正定矩阵 $A\in M_n$ 有秩 n，当且仅当它是正定矩阵. **提示**：考察二次多项式 $p(t)=(x+ty)^*A(x+ty)$，$t\in\mathbf{R}$. 如果 $x^*Ax=0$，证明，对所有 t，$p(t)\geqslant 0$，$p(0)=0$，且在 $t=0$ 时，$dp/dt=0$. 由此推出，对所有 $y\in\mathbf{C}^n$，$y^*Ax=0$，因而 $Ax=0$.

2. 证明，如果半正定矩阵在主对角线上有零元，则它所在的整行和整列必定是零.

3. 证明，如果正定矩阵的主对角元都是 $+1$，则矩阵的所有元的绝对值以 1 为界. 可以等于 1 吗？

4. 证明，半正定矩阵 A 有秩 1，当且仅当对某个非零向量 x，A 具有形式 $A=xx^*$.

5. 设 $A=[a_{ij}]\in M_n$ 是正定矩阵. 证明，矩阵$[a_{ij}/(a_{ii}a_{jj})^{1/2}]$是正定矩阵，它的所有主对角元都是 $+1$，且它的所有元的绝对值以 1 为界. 这样的矩阵称为*相关矩阵*. **提示**：求一个经某个实对角矩阵的相合.

6. 如果 A 有实元素，证明，对所有非零 $x\in\mathbf{R}^n$，要求 $x^TAx>0$ 的条件只与 $H(A)$ 有关.

7. 证明，与(7.1.2)、(7.1.3)、(7.1.4)和(7.1.6)类似的各个论断对使得 $H(A)$ 是正定矩阵的矩阵 $A\in M_n(\mathbf{C})$ 成立.

8. 设 f：$\mathbf{R}\to\mathbf{C}$ 是一个函数，如果对所有 $n=1, 2, \cdots$ 和所有选定的诸点 $\{x_1, x_2, \cdots, x_n\}\subset\mathbf{R}$，矩阵 $[f(x_i-x_j)]\in M_n$ 是半正定矩阵，则称 f 是*正定函数*. 证明，对所有 $x\in\mathbf{R}$，$f(-x)=\overline{f(x)}$. 利用半正定矩阵的行列式非负这一事实证明，如果 f 是正定函数，则

(a) $f(0)\geqslant 0$，　　$n=1$；

(b) f 是有界函数，且对所有 $x\in\mathbf{R}$，$|f(x)|\leqslant f(0)$，　　$n=2$；

(c) 如果 f 在 0 点连续，则 f 处处连续，　　$n=3$.

9. 如果 $f_1(x)$，$f_2(x)$，$\cdots$，$f_n(x)$ 是正定函数，a_1，a_2，$\cdots$，a_n 是非负实数，证明函数 $f(x)=a_1f_1(x)+\cdots+a_nf_n(x)$ 是正定函数.

[400]

10. 证明，对两个给定的 $t\in\mathbf{R}$，函数 e^{itx} 是正定函数. 试用习题 9 证明，对任意选定的诸点 t_1，$\cdots$，$t_n\in\mathbf{R}$ 和任意非负数 a_1，$\cdots$，a_n，$f(x)=a_1e^{it_1x}+\cdots+a_ne^{it_nx}$ 是正定函数.

11. 证明函数 $\cos(x)$ 是正定函数. **提示**：$\cos(x)=(e^{ix}+e^{-ix})/2$.

12. $\sin(x)$ 是正定函数吗？

13. 如果 $g(x)$ 是 $\mathbf{R}$ 上的非负可积函数，证明函数

$$f(x)=\int_{-\infty}^{\infty}e^{ixt}g(t)\mathrm{d}t$$

是正定函数. **提示**：利用定义.

14. 证明函数 $f(x)=1/(1+ix)$ 是正定函数. 提示：在习题 13 中，对 $t>0$，设 $g(t)=e^{-t}$，对 $t\leqslant 0$，设 $g(t)=0$.

15. 由定理(7.5.3)[另见 7.5 节习题 2]可知，如果 $f(x)$ 和 $g(x)$ 都是正定函数，则 $f(x)g(x)$ 亦是正定函数. 证明，如果 $f(x)$ 是正定函数，则 $\overline{f}(x)$ 和 $|f(x)|^2$ 亦是正定函数，然后利用后一个结论由习题 14 推出函数 $1/(1+x^2)$ 是正定函数.

16. 利用(7.0.2)，(7.0.2)以及 $f(x)\equiv 1$ 证明，对所有 $n=1, 2, \cdots$，矩阵 $A=[a_{ij}]\in M_n$（其中，$a_{ij}=1/(i+j-1)$，$i, j=1, 2, \cdots, n$）是正定矩阵.

17. 证明，对所有 $n=1, 2, \cdots$. 矩阵 $A=[a_{ij}]\in M_n$（其中，$a_{ij}=1/(i+j)$，$i, j=1, 2, \cdots, n$）是正定矩阵. **提示**：对所有 $x=[x_i]\in\mathbf{R}^n$，

$$\int_0^{\infty}\left(\sum_{k=1}^{n}x_ke^{-kt}\right)^2\mathrm{d}t\geqslant 0.$$

计算这个积分.

18. 利用(7.1.6)证明，矩阵 $A=[a_{ij}]\in M_n$（其中 $a_{ij}=\min\{i, j\}$）是正定矩阵. **提示**：对 $n=4$，这个矩阵是什么？考察相合 C^TAC，其中 C 是实矩阵，

$$C=\begin{bmatrix}1 & -1 & -1 & \cdots & & -1\\ 0 & 1 & 0 & \cdots & & 0\\ & & \ddots & \ddots & & \\ \vdots & \vdots & & \ddots & \ddots & \vdots\\ & & & & \ddots & 0\\ 0 & 0 & \cdots & & & 1\end{bmatrix}\in M_n.$$

[401] C 为什么是非奇异的？注意这个相合相当于从 A 所有其余的行(列)减去第一行(列). 现在知道了在 C^*AC 的右下方 $(n-1)\times(n-1)$ 子矩阵的形式，再用同样的方式对它实施一个适当的相合来化简这个矩阵. 由此得出 A 相合于 I.

19. 利用习题 18 和极限证法证明，对任意 $N>0$，核 $K(s, t)=\min\{s, t\}$ 在 $[0, N]$ 上是半正定的，也就是说，对于 $[0, N]$ 上的所有连续复值函数 $f(\cdot)$，有

$$\int_0^N\int_0^N K(s,t)\overline{f}(s)f(t)\mathrm{d}s\mathrm{d}t \geqslant 0. \tag{7.1.8}$$

提示：用等距点划分 $[0, N]$，然后把这个积分表示成各个划分上的 Riemann 和的极限.

20. 证明恒等式

$$\int_0^N\int_0^N \min\{s,t\}\overline{f}(s)f(t)\mathrm{d}s\mathrm{d}t = \int_0^N\left|\int_t^N f(s)\mathrm{d}s\right|^2\mathrm{d}t$$

对 $[0, N]$ 上的所有复值连续函数 $f(\cdot)$ 成立，然后用它给出习题 19 中的论断的另一个证法. 这个证法给出 $K(s, t)=\min\{s, t\}$ 是正定的较强的结果；即(7.1.8)中等式成立，当且仅当 $f(t)=0$. **提示**：把二重积分表示成累次积分，然后用分部积分法.

7.2 正定矩阵的特征

正定矩阵有一些有用的和简单的特征.

7.2.1 定理 Hermite 矩阵 $A\in M_n$ 是半正定的，当且仅当它的所有特征值都是非负的. 它是正定的，当且仅当它的所有特征值都是正的.

证明：如果 A 的每个特征值都是正的，则对于任意非零 $x\in\mathbf{C}^n$ 有

$$x^*Ax = x^*U^*DUx = y^*Dy = \sum_{i=1}^n d_i\overline{y}_iy_i = \sum_{i=1}^n d_i\,|\,y_i\,|^2 > 0,$$

其中，$D=\mathrm{diag}(d_1, d_2, \cdots, d_n)$ 是由 A 的特征值组成的对角矩阵，$y=Ux$，且 U 是酉矩阵. 其逆命题包含在论断(7.1.4)，而半正定的情形是类似的. □

[402] **练习** 证明，非奇异矩阵 $A\in M_n$ 是正定的当且仅当 A^{-1} 是正定的.

练习 设 $A\in M_n$ 是半正定矩阵. 试用(7.2.1)证明，A 是正定的，当且仅当 $\mathrm{rank}\,A=n$. 试与(7.1)节习题 1 比较.

7.2.2 推论 如果 $A\in M_n$ 是半正定矩阵，则对所有 $k=1, 2, \cdots$，A^k 也是半正定矩阵.

证明：如果 A 的特征值是 $\lambda_1, \cdots, \lambda_k$，则 A^k 的特征值是 $\lambda_1^k, \cdots, \lambda_n^k$. □

7.2.3 推论 如果 $A=[a_{ij}]\in M_n$ 是 Hermite 矩阵，且是严格对角占优的，又如果对所有 $i=1, 2, \cdots, n$ 有 $a_{ii}>0$，则 A 是正定矩阵.

证明：这是定理(6.1.10)的一部分. 诸条件指出 A 的每个 Geršgorin 圆盘位于开右半平面中. 因 Hermite 矩阵的特征值都是实数，所以 A 的诸特征值必须都是正的，因此由定理(7.2.1)可知 A 是正定矩阵. □

练习 如果 Hermite 矩阵 A 相合于具有正对角元的严格对角占优矩阵，证明 A 是正定矩阵.

下面的特征对于通过计算来确定正定性没有多大的实用价值，但是在理论上它可能很有用.

7.2.4 推论 设 A 是 Hermite 矩阵，又设

$$p_A(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_{n-m}t^{n-m}$$

是 A 的特征多项式. 假定 $0 \leqslant m \leqslant n$ 且 $a_{n-m} \neq 0$. 则 A 是半正定矩阵，当且仅当 $a_k \neq 0$ 对所有 $n-m \leqslant k \leqslant n$ 成立且 $a_k a_{k-1} < 0$ 对 $k=n-m, \cdots, n-1$ 成立. 我们规定 $a_n \equiv 1$.

证明： 论断只是要求，前面的诸系数 a_k 是非零的，且它们的符号是严格交错的. 如果这个条件被满足，$p_A(t)$ 就不可能有任何负的零点；因此 A 的所有特征值必须是非负的. 反过来，如果 A 是半正定矩阵，它的正特征值记作 $\lambda_1, \lambda_2, \cdots, \lambda_m$（其余 $n-m$ 个特征值都是零）. 用归纳法可以证明，各个多项式 $(t-\lambda_1)$，$(t-\lambda_1)(t-\lambda_2)$，$\cdots$，$(t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_m)$ 的诸系数都是非零的，且它们的符号是交错的. 乘以 t^{n-m} 便得 $p_A(t)$. □

为了使下面的特征更容易被接受，用 A_i 表示由 A 前 i 行和前 i 列确定的 A 的前主子矩阵，$A_i \equiv A(\{1, 2, \cdots, i\})$，$i=2, \cdots, n$. 我们已经知道，如果 A 是正定矩阵，则 A 的所有主子式都是正数，事实上，当 A 是 Hermite 矩阵时，逆命题成立. 但是可以得出一个更强的结论. 需要指出的是，如果 A 是 Hermite 矩阵，则每个 A_i 也是 Hermite 矩阵，因此每个 A_i 有实行列式.

403

7.2.5 定理 如果 $A \in M_n$ 是 Hermite 矩阵，则 A 是正定矩阵，当且仅当 $\det A_i > 0$ 对 $i=1, 2, \cdots, n$ 成立. 更一般地，A 的 n 个主子式（不一定是诸前主子式）所组成的任一套序列的正性是 A 为正定矩阵的必要充分条件.

证明： 由(7.1.5)可知，只要 A 是正定矩阵，$\det A_i > 0$ 就对所有 $i=1, 2, \cdots, n$ 都成立. 我们用归纳法和 Hermite 矩阵的交错不等式组(4.3.8)来证明其逆命题，因为 $\det A_1 > 0$，且 A_1 是 1×1 阶的，所以 A_1 是正定矩阵. 如果对某个 $k<n$，A_k 是正定矩阵，则 A_k 的所有特征值都是正数，因而由交错不等式组可知，或许除了 A_{k+1} 的最小特征值以外，A_{k+1} 的所有特征值都是正的. 但是 A_{k+1} 的各特征值之积正好是 $\det A_{k+1}$，根据假定，它是正数，因此 A_{k+1} 不可能有负特征值. 由此可知，A_{k+1} 的最小特征值也是正数，因而 A_{k+1} 必须是正定矩阵. 因为 $A_n = A$，所以 A 是正定矩阵. 对于一般套序列情形，只要考虑 A 的各行和各列的适当置换就可以了. □

定理(7.2.5)说明，当(且仅当)Hermite 的各前主子式是正数的时候，它就是正定矩阵. 再想到(7.2.1)，于是为了验证正定性，可以检验与 A 相关的这两组数中的任何一组.

练习 试用(7.2.5)证明矩阵

$$A = \begin{bmatrix} 5 & -1 & 3 \\ -1 & 2 & -2 \\ 3 & -2 & 3 \end{bmatrix}$$

是正定的.

练习 证明对称矩阵 $\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$ 的各前主子式是非负的，但它不是半正定的.

练习　设 $A\in M_n$ 是 Hermite 矩阵，又假定 $\det A_1>0$，$\det A_2>0$，…，$\det A_{n-1}>0$，且 $\det A_n\geqslant 0$. 证明 A 是半正定矩阵. **提示**：如果将 A_n 的诸特征值与 A_{n-1} 的诸特征值比较，交错不等式指的是什么？

404

练习　假定 Hermite 矩阵 $A\in M_n$ 有全部正对角元和正行列式，考察矩阵

$$\begin{bmatrix}1&2&1\\2&1&1\\1&1&t\end{bmatrix},$$

对适当的 t 值证明，仅有上述假定还不能确定 A 的正定性. 证明，若另有某个 $(n-1)\times(n-1)$ 主子矩阵是对角占优的，则这个假定条件是充分的.

练习　设 $A\in M_n$ 是 Hermite 矩阵，证明，A 是半正定矩阵，当且仅当存在一系列 Hermite 矩阵 A_ε，使得当 $\varepsilon\to 0$ 时 $A_\varepsilon\to A$，且 A_ε 的每个主子矩阵有正行列式. 由此得出，如果 A 的所有主子式都是非负的，则 A 是半正定矩阵.

对所有 $k=1$，2，…，每个正实数有唯一的正的 k 次方根. 类似的结果对正定矩阵也成立.

7.2.6 定理　设 $A\in M_n$ 是半正定矩阵，且 $k\geqslant 1$ 是给定的整数，则存在唯一的半正定 Hermite 矩阵 B 使得 $B^k=A$. 同时还有

(a) $BA=AB$；

(b) $\operatorname{rank} B=\operatorname{rank} A$，因而，只要 A 是正定矩阵，B 就是正定矩阵；

(c) 如果 A 是实矩阵，则 B 也是实矩阵.

证明：我们知道 Hermite 矩阵 A 可酉对角化成 $A=U\Lambda U^*$，其中 $\Lambda=\operatorname{diag}(\lambda_1,\cdots,\lambda_n)$ 且所有 $\lambda_i\geqslant 0$. 定义 $B=U\Lambda^{1/k}U^*$，其中 $\Lambda^{1/k}\equiv\operatorname{diag}(\lambda_1^{1/k},\cdots,\lambda_n^{1/k})$，且每个 λ_i 都取唯一的 k 次非负根. 显然，$B^k=A$ 且 B 是半正定 Hermite 矩阵. 此外，$AB=U\Lambda U^*U\Lambda^{1/k}U^*=U\Lambda\Lambda^{1/k}U^*=U\Lambda^{1/k}\Lambda U^*=U\Lambda^{1/k}U^*U\Lambda U^*=BA$，又因为所有 λ_i(因而它们的 k 次方根)都是非负的，所以 B 是半正定矩阵. B 的秩正好是非零 λ_i 项的个数，它也是 A 的秩. 如果 A 是半正定实矩阵，则我们知道 U 可以选为实正交矩阵，因而在这种情形 B 显然可以选为实矩阵. 余下要考虑的只是唯一性问题.

首先要指出的是，存在多项式 $p(t)$，使得 $p(A)=B$；为了得到 $p(\Lambda)=\Lambda^{1/k}$，从而得到 $p(A)=p(U\Lambda U^*)=Up(\Lambda)U^*=U\Lambda^{1/k}U^*=B$,我们只需选取 $p(t)$ 为适合数组 $(\lambda_1,\lambda_1^{1/k})$，…，$(\lambda_n,\lambda_n^{1/k})$ 的 Lagrange 插值多项式(0.9.11). 另一方面，如果 C 是使得 $C^k=A$ 的任一半正定 Hermite 矩阵，则有 $B=p(A)=p(C^k)$，因而 $CB=Cp(C^k)=p(C^k)C=BC$. 因为 B 和 C 是可交换的 Hermite 矩阵，它们可以同时酉对角化：即存在某个酉矩阵 V 和具有非负对角元的对角矩阵 Λ_1 和 Λ_2，使得 $B=V\Lambda_1V^*$ 和 $C=V\Lambda_2V^*$. 于是从 $B^k=A=C^k$ 的事实可以推出 $\Lambda_1^k=\Lambda_2^k$. 又因为非负数的非负 k 次方根是唯一的，因而得出，$(\Lambda_1^k)^{1/k}=\Lambda_1=\Lambda_2=(\Lambda_2^k)^{1/k}$. 因而 $B=C$.

405

□

上述定理最有用的情形是 $k=2$ 的情形. 正定(半正定)矩阵 A 的唯一正定(半正定)平方根通常记作 $A^{1/2}$. 类似地，对每个 $k=1$，2，…，$A^{1/k}$表示 A 的唯一正定(半正定)k 次方根.

练习 确定$\begin{bmatrix}5 & 3\\3 & 2\end{bmatrix}^{1/2}$.

练习 如果A是正定矩阵，证明$(A^{1/2})^{-1}=(A^{-1})^{1/2}$.

7.2.7 定理 矩阵$B\in M_n$是正定矩阵，当且仅当存在非奇异矩阵$C\in M_n$使得$B=C^*C$.

证明：如果B可以这样表示，则根据(7.1.6)，B是正定矩阵．为了证明可以得到所要求的分解，只要设$C=B^{1/2}$，甚至还可以取C为Hermite矩阵． ┐

7.2.8 推论 Hermite矩阵A是正定矩阵，当且仅当它*相合于单位矩阵.

证明：这只是重述(7.2.7). □

练习 如果$A\in M_n$是正定矩阵，又如果$A=C_1^*C_1$且$A=C_2^*C_2$，其中C_1，$C_2\in M_n$，证明$C_2=VC_1$，其中V是酉矩阵．特别地，证明$A=C^*C$的任何解C具有形式$C=VA^{1/2}$，其中V是酉矩阵．**提示**：证明

$$A^{-1/2}C^*CA^{-1/2}=(CA^{-1/2})^*(CA^{-1/2})=I$$

能够明确指出半正定矩阵A的分解$A=C^*C$，这有时是很有用的．每个方阵C有QR分解(2.6.1)，且C可写成$C=QR$，其中，Q是酉矩阵，而R是与A有相同秩的上三角矩阵．另一方面，$A=C^*C=(QR)^*QR=R^*Q^*QR=R^*R$．如果$C$非奇异，则可以选择$R$使得它的所有对角元都是正数(实际上，存在这种形式的唯一分解$C=QR$)，又如果C为实矩阵，则Q和R均可取为实矩阵．这就证明了下述推论，它给出了A的Cholesky分解． [406]

7.2.9 推论 矩阵A是正定的，当且仅当存在具有正对角元的非奇异下三角矩阵$L\in M_n$使得$A=LL^*$．如果A是实矩阵，则L可以取为实矩阵.

设v_1，…，v_k是内积空间V中k个给定的向量组成的集合，又设$\langle\cdot,\cdot\rangle$是$V$上给定的内积．向量组$v_1$，…，$v_k$关于内积$\langle\cdot,\cdot\rangle$的Gram矩阵是用$g_{ij}=\langle v_j, v_i\rangle$定义的矩阵$G=[g_{ij}]\in M_k$．半正定矩阵的最后一个特征是，它们总是Gram矩阵(7.2.11).

7.2.10 定理 设$G\in M_k$是向量组$\{w_1, \cdots, w_k\}\subset \mathbf{C}^n$关于给定的内积$\langle\cdot,\cdot\rangle$的Gram矩阵，又设$W=[w_1 w_2\cdots w_k]\in M_{n,k}$．则

(a) G是半正定矩阵；

(b) G是非奇异矩阵，当且仅当向量组w_1，…，w_k是无关的；

(c) 存在正定矩阵$A\in M_n$，使得$G=W^*AW$；

(d) rank G=rank W=向量组$\{w_1, \cdots, w_k\}$中极大无关组的向量个数.

证明：如果$G=[g_{ij}]$，且$g_{ij}=\langle w_j, w_i\rangle$，于是因为内积有Hermite性质，所以$G$是Hermite矩阵，且

$$\begin{aligned}x^*Gx&=\sum_{i,j=1}^{k}g_{ij}\bar{x}_ix_j=\sum_{i,j=1}^{k}\langle w_j, w_i\rangle\bar{x}_ix_j=\sum_{i,j=1}^{k}\langle x_jw_j, x_iw_i\rangle\\&=\left\langle\sum_{j=1}^{k}x_jw_j, \sum_{i=1}^{k}x_iw_i\right\rangle=\left\|\sum_{i=1}^{k}x_iw_i\right\|^2\geqslant 0,\end{aligned}$$

其中$\|\cdot\|$是由已给内积诱导的范数．根据范数的正定性，只有当

$$\sum_{i=1}^{k} x_i w_i = 0$$

时等式才成立，而且只有当给定的向量组相关时，上式才对非平凡系数组 x_i 成立．如果 G 是奇异矩阵，则存在某个非零向量 x 使得 $Gx=0$，因而 $x^*Gx=0$，这推出向量组 w_i 是相关的．反过来，如果 $x_1w_1+\cdots+x_kw_k=0$，且 $x=[x_i]\neq 0$，则我们已经证明 $x^*Gx=0$，因而 G 一定是奇异矩阵．

407

如果 $\{e_1, \cdots, e_n\}$ 是 $\mathbf{C}^n$ 的标准正交基，则根据(a)和(b)，$A=(\langle e_j, e_i\rangle)$ 是正定矩阵．对任意向量 x，$y\in\mathbf{C}^n$ 有

$$\langle y,x\rangle = \left\langle \sum_{j=1}^{n} y_j e_j, \sum_{i=1}^{n} x_i e_i \right\rangle = \sum_{i,j=1}^{n} \langle e_j, e_i\rangle \bar{x}_i y_j = x^* A y,$$

于是有 $g_{ij}=\langle w_j, w_i\rangle = w_i^* A w_j$，因而 $G=W^*AW$．

最后，如果 $Gx=0$，则 $x^*Gx=x^*W^*AWx=(Wx)^*A(Wx)=0$，因为 A 是正定矩阵，这就蕴涵 $Wx=0$．反之，$Wx=0$ 蕴涵 $Gx=W^*A(Wx)=0$，于是 A 与 W 有相同的零空间，因而有相同的秩．W 的列秩是向量组 $\{w_1, \cdots, w_k\}$ 中极大无关组的向量个数． □

练习 定理最常见的应用是针对所给内积为普通 Euclid 内积 $\langle x, y\rangle = y^*x$ 的情形的．证明，在这种情形，$A=I$，并且证明给定的向量组 $\{w_1, \cdots, w_k\}\subset\mathbf{C}^n$ 中极大无关组的向量个数恰好是矩阵 $G=[w_i^*w_j]\in M_k$ 的秩．

7.2.11 推论 设 $A\in M_n$ 是给定的矩阵．则 A 是秩为 $r\leqslant n$ 的半正定矩阵，当且仅当存在恰好含有 r 个无关向量的向量组 $S=\{w_1, \cdots, w_n\}\subset\mathbf{C}^n$，使得 A 是 S 关于 Eucldi 内积的 Gram 矩阵．

证明：充分性部分在上述定理中已有论述．至于必要性，可以利用(7.2.6)把 A 写成 $A=B^2$，B 是半正定矩阵．B 的秩与 A 的秩相同，且 $A=B^2=B^*B$ 是 B 的各列在 Euclid 内积下的 Gram 矩阵． □

习题

1. 证明，如果 A 是 Hermite 矩阵，则对所有 $k=1, 2, \cdots$，A^{2k} 是半正定矩阵，而 e^A 是正定矩阵．参看(5.6.15)下面的练习．

2. 如果 A 是半正定矩阵，又如果 $p(t)$ 是使 $p(t)>0$ 时对所有 $t\geqslant 0$ 都成立的任一多项式，证明 $p(A)$ 是半正定矩阵．**提示**：$p(A)$ 的诸特征值是什么？这是如何推广了习题 1？

3. 试用(7.2.5)证明，用 $a_{ij}\equiv\min\{i, j\}$ 定义的矩阵 $A=[a_{ij}]\in M_n$ 是正定矩阵．**提示**：计算 $\det A_i$；从所有其余行中减去第 1 行，然后对第 1 列也这样做，$a_{ij}\equiv\max\{i, j\}$ 说明什么？

408

4. 如果 A 和 B 是正定矩阵，证明直和 $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ 也是正定矩阵．

5. 给出一个(非 Hermite)实方阵的例子，它的各前主子式都是正数，但使某个特征值有负实部．

6. 试给出(7.2.5)中一般不等式组的详细证明．即证明，n 个主子式(不一定是诸前主子式，按包含关系)所构成的任一套序列的正性是 $n\times n$ Hermite 矩阵为正定矩阵的充分条件．

7. 如果用 A 的诸子式的符号来表示，A 是负定(半负定)的必要充分条件是什么？

8. 半正定矩阵 A 有不同于 $A^{1/2}$ 的“平方根”吗？有多少？有不同于 $A^{1/2}$ 的 k 次方根吗？有非 Hermite 平方根吗？**提示**：考察$\begin{bmatrix}-1 & 1\\ 0 & 1\end{bmatrix}^2$.

9. 如果 $B\in M_n$ 是半正定矩阵，且有秩 m，证明，存在秩为 m 的 $m\times n$ 矩阵 C，使得 $B=C^*C$. 特别要指出的是，秩为 1 的半正定矩阵总可以写成形式 xx^*，其中 $x\in\mathbf{C}^n$ 为某个向量.

10. 假定 $A\in M_n$ 是半正定矩阵，且有秩 $r<n$. 证明 A 有 $r\times r$ 阶正定主子矩阵.

11. 设 $A\in M_n$ 是 Hermite 矩阵，证明，A 是正定矩阵，当且仅当经典伴随 adj A 是正定矩阵且 det $A>0$. 如果 A 是半正定矩阵，证明，adj A 是半正定矩阵且 det $A\geqslant 0$. **提示**：考察 $A_\varepsilon\equiv A+\varepsilon I$，$\varepsilon>0$. 试考察 $A\equiv\mathrm{diag}(0,0,-1)$ 来说明，如果 A 不是半正定矩阵，也可能 adj A 是半正定矩阵且 det $A\geqslant 0$.

12. 已知 $r\in(0,1)$，考虑由 $a_{ij}=r^{|i-j|}$ 定义的实对称 Toeplitz 矩阵 $A=[a_{ij}]\in M_n$. 试如下证明 A 是正定矩阵：(a)如果 A_{ij} 是 A 的 i，j 子式，证明，只要 $|i-j|\geqslant 2$ 就有 det $A_{ij}=0$. **提示**：如果 $i=1$ 且 $j>2$，则可以看出 A_{1j} 的第 1 列是第 2 列的倍数. (b)设 $D_n=\det A$. 证明 $D_2=1-r^2$，然后将 D_{n+1} 按第 1 行的余子式展开，且利用(a)证明 $D_{n+1}=D_n-r^2D_n=(1-r^2)D_n=(1-r^2)^n$. (c)利用(7.2.5)推出 A 是正定矩阵.

13. 证明习题 12 中的矩阵 A 有一个实对称三对角矩阵为其逆，再证明，$(1-r^2)A^{-1}$ 在上对角线和下对角线的每个位置上有元素 $-r$，且它有主对角元 1，$1+r^2$，…，$1+r^2$，1. **提示**：利用习题 12(a)证明 A^{-1} 是三对角矩阵. 为什么 A^{-1} 一定是对称矩阵？然后利用 $AA^{-1}=A^{-1}A=I$ 确定 A^{-1} 的各元素.

409

14. 设 $\langle\cdot,\cdot\rangle$ 是 $\mathbf{C}^n$ 上给定的内积，设 $\mathscr{B}=\{e_1,\cdots,e_n\}$ 是 $\mathbf{C}^n$(关于普通的 Euclid 内积)的标准正交基，又设 $G\in M_n$ 表示 $\mathscr{B}$ 关于所给内积 $\langle\cdot,\cdot\rangle$ 的 Gram 矩阵. 证明，对所有 x，$y\in\mathbf{C}^n$ 有

$$\langle x,y\rangle=y^*Gx. \tag{7.2.12}$$

由此可得，函数 $\langle\cdot,\cdot\rangle$：$\mathbf{C}^n\times\mathbf{C}^n\to\mathbf{C}$ 是内积，当且仅当存在正定矩阵 G 使得(7.2.12)成立.

15. 回忆一下在(5.4.12)中定义的对偶范数概念. 设 $\langle\cdot,\cdot\rangle$ 是 $\mathbf{C}^n$ 上给定的内积，且设 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上给定的范数. 所给范数不一定是由所给内积诱导的. 我们可以定义 $\|\cdot\|$ 关于内积 $\langle\cdot,\cdot\rangle$ 的对偶范数为

$$\|x\|^D_{\langle\cdot,\cdot\rangle}\equiv\max_{\|y\|=1}|\langle x,y\rangle|.$$

注意，如果 $\langle\cdot,\cdot\rangle$ 是普通的 Euclid 内积，则它就是 $\|\cdot\|$ 的普通的对偶范数. 对偶范数概念的这种推广可以产生用其他方法尚未得到过的向量范数吗？**提示**：利用习题 14，记 $\langle x,y\rangle=y^*Gx$，然后证明

$$\|x\|^D_{\langle\cdot,\cdot\rangle}=\|G^{-1}x\|^D\equiv(\|x\|_{G^{-1}})^D$$

16. 设 $A\in M_n$ 是给定的矩阵，证明，$\rho(A)<1$，当且仅当存在正定矩阵 $B\in M_n$ 使得 $B-A^*BA$ 是正定矩阵. **提示**：如果 B 是正定矩阵，设 $C=B^{1/2}$. 如果

$$B-A^*BA=C^*C-(CA)^*(CA)$$

是正定矩阵，则对任何非零 $x\in\mathbf{C}^n$ 有

$$x^*[C^*C-(CA)^*(CA)]x>0,$$

或 $\|Cx\|_2>\|CAx\|_2$. 设 $y=Cx$. 证明对所有非零 $y\in\mathbf{C}^n$ 有 $\|y\|_2>\|CAC^{-1}y\|_2$，由此得出 $\|CAC^{-1}\|_2<1$. 因此 $\rho(A)=\rho(CAC^{-1})\leqslant\|CAC^{-1}\|_2<1$. 反过来，如果 $\rho(A)<1$，则存在非奇异矩阵 $C\in M_n$ 使得 $\|CAC^{-1}\|<1$[见(5.6)节习题25]，并且上述证明可以反推回去，再令 $B=C^*C$.

17. 设 A，$B\in M_n$ 是半正定矩阵，且不都是奇异矩阵. 证明 $\|A-B\|_2\leqslant\|A^2-B^2\|_2/[\lambda_{\min}(A)+\lambda_{\min}(B)]$. **提示**：设 $E=A-B$. 设 $x\in\mathbf{C}^n$ 是使 $Ex=\lambda x$ 且 $|\lambda|=\rho(E)=\|E\|_2$ 的单位向量. 于是，$A^2-B^2=AE+EA-E^2$，并且 $\|A^2-B^2\|_2\geqslant|x^*(AE+EA-E^2)x|=|\lambda|(x^*Ax+x^*Bx)\geqslant|\lambda|(\lambda_{\min}(A)+\lambda_{\min}(B))$.

410

18. 设 A、$B\in M_n$ 是半正定矩阵，且假定 A 是正定矩阵. 利用习题17证明

$$\|A^{1/2}-B^{1/2}\|_2\leqslant\|A^{-1/2}\|_2\|A-B\|_2 \tag{7.2.13}$$

并且说明为什么这个不等式蕴涵以下事实：定义在由 M_n 中的半正定矩阵所组成的集合上的函数 $f: C\to C^{1/2}$ 在这个集合的内部(它是由正定矩阵组成的开集)连续. 写出并直接证明关于 $[0,\infty]$ 上的普通纯量的平方根函数 $f: t\to\sqrt{t}$ 的不等式，这个不等式是从(7.2.13)中令 $n=1$ 得来的.

7.3 极形式和奇异值分解

下面，论述(不一定是方阵的)复矩阵的两种相关的重要分解，它们与正定性概念有密切关系.

7.3.1 引理 设 $A\in M_{m,n}$，$m\leqslant n$，且 rank $A=k\leqslant m$. 则存在酉矩阵 $X\in M_m$、具有非负对角元 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k>\lambda_{k+1}=\cdots=\lambda_m=0$ 的对角矩阵 $\Lambda\in M_m$、以及具有标准正交行的 $Y\in M_{m,n}$，使得 $A=X\Lambda Y$. 矩阵 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_m)$ 总是唯一确定的，且 $\{\lambda_1^2,\cdots,\lambda_m^2\}$ 是 AA^* 的诸特征值. 矩阵 X 的各列是 AA^* 的特征向量. 如果 AA^* 有互不相同的特征值，那么，X 可确定到相差一个右对角因子 $D=\mathrm{diag}(e^{i\theta_1},\cdots,e^{i\theta_m})$，其中所有 $\theta_i\in\mathbf{R}$；也就是说，如果 $X_1\Lambda Y_1=X_2\Lambda Y_2$，则 $X_2=X_1D$. 给定 X 后，如果 rank $A=m$，则矩阵 Y 是唯一确定的. 如果 A 是实矩阵，则 X 和 Y 可以取实矩阵.

证明：如果 $A=X\Lambda Y$ 是所要求的分解形式，则 $AA^*=X\Lambda YY^*\Lambda X^*=X\Lambda I\Lambda X^*=X\Lambda^2X^*$，因而 $X\Lambda^2X^*$ 是 Hermite 矩阵 AA^* 的酉对角化. 若 $X=[x_1x_2\cdots x_m]$ 且 $\Lambda^2=\mathrm{diag}(\lambda_1^2,\cdots,\lambda_m^2)$，则 $AA^*x_j=\lambda_j^2x_j$，$j=1,2,\cdots,m$，且向量组 $\{x_j\}$ 是标准正交组. 因为 Λ 的诸对角元是非负的，且按递减顺序排列，所以 Λ 由 AA^* 唯一确定. 如果数组 $\{\lambda_i^2\}$ 是互不相同的，那么，除了相差一个模为1的复纯量因子以外，AA^* 的各个相应的正规化特征向量都是确定的，因此，如果 X_1 和 X_2 是其各列为 AA^* 的特征向量的两个酉矩阵，则一定有 $X_2=X_1D$，其中，$D=\mathrm{diag}(d_1,\cdots,d_n)$，且所有 $|d_i|=1$.

AA^* 的相应于重特征值的特征向量不是唯一确定的，但是，它们一经选定，且把它们标准正交化，酉矩阵 X 就固定下来. 如果 Λ 是非奇异矩阵，也就是当 $k=\mathrm{rank}\ A=m$ 的时候，则 Y

$=\Lambda^{-1}X^*A$ 是唯一确定的. 我们不难验证, $YY^*=\Lambda^{-1}X^*(AA^*X)\Lambda^{-1}=\Lambda^{-1}X^*X\Lambda^2\Lambda^{-1}=\Lambda^{-1}\Lambda^2\Lambda^{-1}=I$, 因此, 这个矩阵 Y 有标准正交行. [411]

余下只要讨论 rank $A=k<m$ 的情形. 因为当所有 $\lambda_i\neq 0$ 时, 我们希望有 $Y=\Lambda^{-1}X^*A=\Lambda^{-1}(A^*X)^*$, 这促使我们定义 Y 的第 j 行是行向量 y_j^*, 其中 $y_j=\lambda_j^{-1}(A^*x_j)$, $j=1,\cdots,k$. 于是

$$[\lambda_j^{-1}(A^*x_j)]^*[\lambda_k^{-1}(A^*x_k)]=x_j^*AA^*x_k/\lambda_j\lambda_k=x_j^*\lambda_k^2x_k/\lambda_j\lambda_k=x_j^*x_k\lambda_k/\lambda_j$$

如果 $j\neq k$, 它就是零, 如果 $j=k$, 它就是 1, 这是因为向量组 $\{x_j\}$ 是标准正交组. 向量组 $\{y_1,\cdots,y_k\}$ 是 $\mathbf{C}^n$ 中的标准正交组, 且 $n\geqslant m>k$, 因此另外有 $m-k$ 个(但不是唯一确定的)标准正交向量 $y_{k+1},\cdots,y_m$, 使得矩阵 $Y^*=[y_1y_2\cdots y_ky_{k+1}\cdots y_m]\in M_{n,m}$ 有 m 个标准正交列.

现在证明 $X^*A=\Lambda Y$. 根据向量 y_j 的定义, 这个恒等式两边的前 k 行是相等的. 因为 Λ 的最后 $m-k$ 个对角元是零, 所以右边最后 $m-k$ 行都是零; 左边最后 $m-k$ 行也都是零, 这是因为, 如果 $AA^*x_j=0$, 则 $0=x_j^*AA^*x_j=(A^*x_j)^*(A^*x_j)=0$, 因而 $A^*x_j=0$.

最后, 如果 A 是实矩阵, 则 AA^* 也是实矩阵且有实特征值, 因而, 特征向量组 X 可以取为实向量组. 根据定义, 由 X 确定的 Y 的前 k 个行都是实的, 添加的 $m-k$ 个正交单位向量可以取实的. 因此, 如果 A 是实矩阵, 它的所有因子都可以取实矩阵. □

每个非零复数 j 有唯一的"极表示"$z=pu$, 其中 p 是正实数, 而 u 是模为 1 的复数. 实际上, 如果 $z\neq 0$, 则 $p=|z|$ 且 $u=p^{-1}z=z/|z|$. 如果 $z=0$, 则 z 仍然可以写成 $p=0$ 的极形式, 不过 u 不再是唯一确定的. 当然, u 可以取模为 1 的任意复数.

如何把这种极表示推广到复矩阵 $A\in M_n$ 呢? 一种回答是 $A=PU$, 其中, P 是正定(半正定)矩阵, 而 U 是酉矩阵. 甚至还可以把它推广到 A 不是方阵的情形.

7.3.2 定理 设 $A\in M_{m,n}$, 且 $m\leqslant n$. 则 A 可以写成

$$A=PU,$$

其中, $P\in M_m$ 是半正定矩阵, rank $P=$rank A, 而 $U\in M_{m,n}$ 有标准正交行(即 $UU^*=I$). 矩阵 P 总可以唯一确定为 $P=(AA^*)^{1/2}$, 而当 A 有秩 m 时, U 是唯一确定的. 如果 A 是实矩阵, 则 P 和 U 可以取实矩阵. [412]

证明: 利用(7.3.1)把 A 写成 $A=X\Lambda Y=X\Lambda X^*XY$, 且令 $P=X\Lambda X^*$ 和 $U=XY$. 于是 P 是半正定矩阵, 且 $UU^*=XYY^*X^*=XIX^*=XX^*=I$, 因而 U 有标准正交行. 根据(7.3.1)中的构造法, $P=(AA^*)^{1/2}$, 一般说来, 如果 $A=PU$, 则 $AA^*=PUU^*P=P^2$, 所以 P 一定总是 AA^* 的(唯一)半正定方根. 如果 A 有秩 m, 则 P 非奇异, 而且 $U=P^{-1}A$ 是唯一确定的. 但是, 正如在(7.3.1)中看到的, 如果 rank $A<m$, 则 Y 的相应于 P 的 0 特征值的各行不是唯一确定的, 因此, 当 rank $A<m$ 时, $U=XY$ 未必是唯一确定的. □

本定理直接推出下述重要的特殊情形.

7.3.3 推论 如果 $A\in M_n$, 则它可以写成形式

$$A=PU,$$

其中, P 是半正定矩阵, U 是酉矩阵. 矩阵 P 总可以唯一确定为 $P\equiv(AA^*)^{1/2}$; 如果 A 是非奇异矩阵, 则 U 可以唯一确定为 $U\equiv P^{-1}A$. 如果 A 是实矩阵, 则 P 和 U 可以取实矩阵.

练习 说明定理(7.3.2)可以用下述极限证法来证明. 如果 A 是非奇异矩阵，则令 $P\equiv(AA^*)^{1/2}$，定义 $U=P^{-1}A$，且验证 $UU^*=I$. 因此，P 和 U 都是唯一确定的. 如果 A 是奇异矩阵，考虑 $A_\varepsilon\equiv A+\varepsilon I$，以及形式 $A_\varepsilon=P_\varepsilon U_\varepsilon$，其中两个因子都是唯一确定的. 利用选择原理(2.1.8)得到序列 $\varepsilon_k\to 0$(当 $k\to\infty$ 时)，使得当 $k\to\infty$ 时，U_{ε_k} 按对应元收敛于酉矩阵 U. 因为 $P_{\varepsilon_k}=A_{\varepsilon_k}U_{\varepsilon_k}^*$，我们还有 $P_{\varepsilon_k}\to P$ 和 $A=PU$. 值得指出的是，从理论上讲，这个证明虽然比前面对(7.3.2)所做的证明更简略，但是当 A 是奇异矩阵时，它并没有给出得到因子 P 和 U 的构造性方法.

分解(7.3.2)称为矩阵 A 的*极形式*或*极分解*. 当 A 满秩时，两个因子还是唯一的.

练习 如果 $A\in M_{m,n}$，且 $m\geqslant n$，证明它可以写成

$$A=WQ,$$

其中，$W\in M_{m,n}$ 有标准正交列(即 $W^*W=I$)，而 $Q\in M_n$ 是半正定矩阵. **提示**：利用(7.3.2)分解 A^*.

练习 设 $x\in\mathbf{C}^n$ 是给定的非零向量，且 $A\equiv x\in M_{n,1}$. 证明 A 的极分解是 $A=x=\|x\|_2u$，其中 $u\equiv x/\|x\|_2$. 故极分解可以看作非零向量的简便分解 $x=\|x\|_2(x/\|x\|_2)$ 到矩阵的推广.

[413]

练习 证明，方阵 A 既可写成 $A=PU$，也可写成 $A=WQ$，其中，$P=(AA^*)^{1/2}$，而 $Q=(A^*A)^{1/2}$. 有时称它们为 A 的"左"和"右"极分解. 证明，唯一确定的半正定因子 P 和 Q 相等，当且仅当 A 是正规矩阵. 实际上，如果 A 是非奇异矩阵，则唯一确定的酉因子 U 和 W 恒相等[定理(7.3.6)前的练习].

练习 不是每个方阵都是正规的；即 $AA^*=A^*A$ 未必成立. 但是 AA^* 总酉相似于 A^*A. 试用极分解(7.3.3)证明这一事实.

7.3.4 定理 设 $A\in M_n$，又设 $A=PU$ 是极分解. 则 A 是正规矩阵，当且仅当 $PU=UP$.

证明：如果 P 和 U 可交换，则 $AA^*=PUU^*P^*=PP=P^2$，$A^*A=U^*P^*PU=U^*P^2U=U^*UP^2=P^2$，因而 A 是正规矩阵. 如果 A 是正规矩阵，则 $P^2=U^*P^2U$. 我们知道，P^2 和 U^*P^2U 都是半正定方阵，且显然有相应的半正定平方根 P 和 U^*PU. 但是定理(7.2.6)说明，这样的平方根是唯一的，因而 $P=U^*PU$ 或 $UP=PU$. □

我们的下一个目标是，从(7.3.1)得出(不一定是方阵的)任意矩阵的奇异值分解.

7.3.5 定理 如果 $A\in M_{m,n}$有秩 k，则它可以写成形式

$$A=V\Sigma W^*,$$

其中 $V\in M_m$ 和 $W\in M_n$ 是酉矩阵. 矩阵 $\Sigma=[\sigma_{ij}]\in M_{m,n}$对所有 $i\neq j$ 有 $\sigma_{ij}=0$，且 $\sigma_{11}\geqslant\sigma_{22}\geqslant\cdots\geqslant\sigma_{kk}>\sigma_{k+1,k+1}=\cdots=\sigma_{qq}=0$，其中 $q=\min\{m,n\}$. 数 $\{\sigma_{ii}\}\equiv\{\sigma_i\}$是 AA^* 的特征值的非负平方根，因而被唯一确定. V 的各列是 AA^* 的特征向量，W 的各列是 A^*A 的特征向量(它们的排列顺序与相应的特征值 σ_i^2 的排列顺序相同). 如果 $m\leqslant n$，且 AA^* 有互不相同的特征值，则 V 可以确定到相差一个右对角因子 $D=\mathrm{diag}(e^{i\theta_1},\cdots,e^{i\theta_n})$，其中所有 $\theta_i\in\mathbf{R}$；也就是，如果 $A=V_1\Sigma W_1^*=V_2\Sigma W_2^*$，则 $V_2=V_1D$. 如果 $m<n$，则 W 一定不是唯一确定的；如果 $n=m$，且 V 已经给定，那么 W 是唯一确定的. 如果 $n\leqslant m$，则 V 和 W 的唯一性要根据 A^* 来确定. 如果 A

[414]

是实矩阵，则 V，Σ 和 W 都可以取实矩阵.

证明： 不失一般性，假定 $m\leqslant n$（否则用 A^* 代替 A）. 利用(7.3.1)可把 A 写成 $A=X\Lambda Y$，其中，X，$\Lambda\in M_m$，且 $Y\in M_{m,n}$. 令 $V\equiv X$，取 $\Sigma\equiv[\Lambda\,|\,0]\in M_{m,n}$，且定义 $W\equiv[Y^*\,|\,S^*]\in M_n$ 使得 W 的各列是 $\mathbf{C}^n$ 中的标准正交组. Y^* 的各列已是标准正交组，所以，如果 $m<n$，可选取 $S^*\in M_{n,(n-m)}$ 的各列（但不唯一）使 W 为酉矩阵. 这直接推出 $V\Sigma W^*=X\Lambda Y=A$. 关于唯一性的论断可以从(7.3.1)中的相应论断得出. □

对 $i=1,\cdots,q=\min\{m,n\}$，Σ 的"对角元"$\sigma_i=\sigma_{ii}$ 称为 $A\in M_{m,n}$ 的奇异值（有时只对非零对角元才这样称呼），而 V 的列和 W 的列是 A 的（相应左和右）奇异向量. 分解(7.3.5)称为 A 的奇异值分解. 极矩阵 P 是 AA^* 的唯一半正定平方根，而奇异值 σ_i 是 AA^* 的特征值的非负平方根，所以 A 的奇异值与极矩阵 P 的特征值相同. 虽然奇异值按递减顺序排列是方便的，但这并不是奇异值分解中的一般规定；它只是由 A 唯一确定的一组奇异值.

应该指出，奇异值分解是正规矩阵的酉对角化到任意矩阵的自然推广. 由于这个理由，常常有这种情形，关于正规矩阵的特征值的一些结果可以推广成关于一般矩阵的奇异值的一些结论.

练习 设 $x\in\mathbf{C}^n$ 是给定的非零向量，且 $A\equiv x\in M_{n,1}$. 证明，A 的奇异值分解是 $A=x=V\Sigma W^*$，其中，$W=[1]\in M_1$，$\Sigma=[\|x\|_2,0,\cdots,0]^T\in M_{n,1}$，而 $V=[v_1,\cdots,v_n]\in M_n$ 有 $v_1=x/\|x\|_2$，且 $v_2,\cdots,v_n$ 是与 x 正交的任意 $n-1$ 个单位正交向量.

如果 $A\in M_n$，则奇异值分解中的三个因子 V，Σ 和 W 都是 $n\times n$ 矩阵. 如果 $A=PU$ 是 A 的极分解，又如果 $P=V\Lambda V^*$ 是 P 的酉对角化，其中 P 的（一定非负）诸特征值按递减顺序排列，则 $A=PU=V\Lambda V^*U=(V)(\Lambda)(V^*U)=V\Lambda W^*$ 是 A 的奇异值分解，其中 $V=V$，$\Sigma=\Lambda$，且 $W=U^*V$. 注意到 $AA^*=V\Sigma W^*W\Sigma V^*=V\Sigma^2V^*$，因而 V 的诸列是 Hermite 矩阵 AA^* 的相应于特征值 $\sigma_1^2,\cdots,\sigma_n^2$ 的特征向量. 另外，$A^*A=W\Sigma V^*V\Sigma W^*=W\Sigma^2W^*$，因而 W 的诸列是 A^*A 的特征向量. [415]

练习 如果 $A\in M_n$ 是非奇异矩阵，证明下述步骤给出奇异值分解 $A=V\Sigma W^*$：

(a) 作出正定 Hermite 矩阵 AA^*，然后通过求 AA^* 的各（正）特征值 $\{\lambda_i\}$ 以及相应的正规化特征向量组 u_i 算出酉对角化 $AA^*=U\Lambda U^*$.

(b) 令 $\Sigma=\Lambda^{1/2}$ 和 $V=U=[u_1\cdots u_n]$.

(c) 令 $W\equiv A^*V\Sigma^{-1}$.

证明 W 是酉矩阵且 $A=V\Sigma W^*$. **提示：** 计算 W^*W.

练习 设 $A\in M_n$ 是给定的（不一定非奇异）矩阵，证明下述步骤给出奇异值分解 $A=V\Sigma W^*$：

(a) 存在某个 $c=c(A)>0$，使得对所有正 $\varepsilon<c$，$A_\varepsilon=A+\varepsilon I$ 是非奇异矩阵. 设 $0<\varepsilon<c$.

(b) 利用上一个练习中的方法作奇异值分解 $A_\varepsilon=V_\varepsilon\Sigma_\varepsilon W_\varepsilon^*$.

(c) 利用选择原理(2.1.8)且设 $\varepsilon\to0$，通过值 ε_k 的序列使得

$$\lim_{\varepsilon_k\to0}V_{\varepsilon_k}=V\quad 和\quad \lim_{\varepsilon_k\to0}W_{\varepsilon_k}=W$$

都存在.

(d) 证明 $A=V\Sigma W^*$，其中 $\Sigma=\lim_{\epsilon\to 0}\Sigma_\epsilon$.

这个可以用来证明一般的奇异值分解(7.3.5)的证法保证存在一个奇异值分解，但是，当 A 不是满秩的时候，它一般没有给出计算奇异值分解中的各个因子的构造性方法.

练习 假定 $A\in M_n$ 是非奇异矩阵，且 $A=PU$，$A=WQ$ 是 A 的左和右极分解，其中 P，$Q\in M_n$ 是正定矩阵，U，$W\in M_n$ 是酉矩阵. 证明恒有 $U=W$，但是，$P=Q$ 当且仅当 A 是正规矩阵. 如果 A 是奇异矩阵，证明存在 A 的左和右的极分解使 $U\neq W$. **提示**：如果 $A=V\Sigma W^*$ 是 A 的奇异值分解，则 V 或 W 都不是唯一确定的，但是，$A=(VW^*)(W\Sigma W^*)=(V\Sigma V^*)(VW^*)$；采用(7.3.3)的唯一性部分. 考察 $A=0$，说明，如果 A 是奇异矩阵，则 A 的两个极分解中的酉因子不一定相同.

[416]

如果 $A\in M_n$ 是正规矩阵，且 $A=V\Sigma W^*$ 是奇异值分解，则 $AA^*=A^*A$，因而 AA^* 和 A^*A 有相同的特征向量. 但是，由此不能推出，在 A 的奇异值分解中有 $V=W$，因为 $V=W$ 时，$A=V\Sigma V^*$ 就一定是(甚至是半正定)Hermite 矩阵. 如果 $A=U\Lambda U^*$ 是 A 的酉对角化，且 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，则每个 $\lambda_k=|\lambda_k|e^{i\theta_k}$ 对某个 $\theta_k\in\mathbf{R}$ 成立；如果 $\lambda_k=0$，选 $\theta_k\equiv 0$. 如果令 $D\equiv\mathrm{diag}(e^{i\theta_1},\cdots,e^{i\theta_n})$ 和 $|\Lambda|\equiv\mathrm{diag}(|\lambda_1|,\cdots,|\lambda_n|)$，则 $\Lambda=|\Lambda|D$，且 $A=U\Lambda U^*=U|\Lambda|DU^*=(U)(|\Lambda|)(U\bar{D})^*\equiv V\Sigma W^*$ 是 A 的奇异值分解，其中，$V=U$，$\Sigma=|\Lambda|$，且 $W=U\bar{D}$.

因此，正规矩阵的奇异值正好是其特征值的绝对值，V 的各列是 A 的特征向量，W 的各列视为与 V 的各列相同，只是每一列都乘以一个由相应的特征值确定的绝对值为 1 的复纯量. 如果 A 是 Hermite 矩阵，则所有特征值都是实的，$D=\bar{D}$ 且 $D=\mathrm{diag}(\mathrm{sgn}(\lambda_1),\cdots,\mathrm{sgn}(\lambda_n))$，其中令 $\mathrm{sgn}(0)=1$. 如果 A 是正定 Hermite 矩阵，则 $D=I$，$V=W=U$，且 $\Lambda=\Sigma$.

Schur 三角化定理(2.3.1)的一个有效应用是，证明了每个复方阵是具有互异特征值的矩阵的极限. 奇异值分解可以用来证明每个复矩阵(方阵或非方阵)是具有互异奇异值的矩阵的极限. 这可能是有用的，因为在奇异值互不相同的情形，奇异值分解具有不完全的唯一性.

7.3.6 推论 如果 $A\in M_{m,n}$ 是给定的矩阵，又 $\|\cdot\|$ 是 $M_{m,n}$ 上的给定的范数，则对每个 $\epsilon>0$，存在具有互异奇异值的 $A_\epsilon\in M_{m,n}$ 使得 $\|A-A_\epsilon\|<\epsilon$.

证明：假定 $m\leqslant n$. 设 $A=V\Sigma W^*$ 是 A 的奇异值分解，又设

$$\Sigma_\delta\equiv[\mathrm{diag}(\sigma_1+\delta,\sigma_2+2\delta,\cdots,\sigma_m+m\delta)\mid 0]$$

其中 $0\in M_{m,n-m}$. 如果 A 的所有奇异值都相等，则对所有 $\delta>0$，Σ_δ 将有互不相同的对角元. 不然的话，如果选取 $\delta>0$ 使得 $m\delta$ 小于各个不同奇异值间的最小的差，则 Σ_δ 就有互不相同的对角元. 在这两种情形，当 $\delta\to 0$ 时，$\Sigma_\delta\to\Sigma$. 如果令 $A_\delta\equiv V\Sigma_\delta W^*$，因为 Frobenius 范数是酉不变的，所以当 $\delta\to 0$ 时，$\|A-A_\delta\|_2=\|\Sigma-\Sigma_\delta\|_2\to 0$. 但是，$M_{m,n}$ 上的所有范数都是等价的，因此完成了证明. 如果 $m>n$，证明是类似的. □

[417]

有一个简单的变换使我们能把 Hermite 矩阵的关于特征值的结果转变成任意矩阵的关于奇异值的结果.

7.3.7 定理 设 $A\in M_{m,n}$，$q=\min\{m,n\}$，并且定义 $\tilde{A}\in M_{m+n}$ 为

$$\tilde{A}\equiv\begin{bmatrix}0 & A\\ A^* & 0\end{bmatrix}.\qquad(7.3.7a)$$

设 $\sigma_1, \sigma_2, \cdots, \sigma_q$ 是非负实数. 则 A 的奇异值是 $\sigma_1, \sigma_2, \cdots, \sigma_q$，当且仅当 $\tilde{A}$ 的 $m+n$ 个特征值是 $\sigma_1, \sigma_2, \cdots, \sigma_q, -\sigma_1, -\sigma_2, \cdots, -\sigma_q$ 和 $|m-n|$ 个 0.

证明： 假定 $m\geqslant n$，且设 $A=V\Sigma W^*$ 是 A 的奇异值分解. 记

$$\Sigma=\begin{bmatrix}S\\0\end{bmatrix}\in M_{m,n},\quad 0\in M_{m-n,n},$$

其中 $S=\mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$，并且把酉因子 $V\in M_m$ 写成 $V=[V_1\,|\,V_2]$，其中 $V_1\in M_{m,n}$，$V_2\in M_{m,(m-n)}$. 如果令 $\hat{V}\equiv V_1/\sqrt{2}$ 和 $\hat{W}\equiv W/\sqrt{2}$，则矩阵

$$U\equiv\begin{bmatrix}\hat{V} & -\hat{V} & V_2\\ \hat{W} & \hat{W} & 0\end{bmatrix}\in M_{m+n},\quad 0\in M_{n,m-n}$$

是酉矩阵，且经直接计算可以验证

$$\tilde{A}=U\begin{bmatrix}S & 0 & 0\\ 0 & -S & 0\\ 0 & 0 & 0\end{bmatrix}U^*,$$

其中对角零块是 $(m-n)\times(m-n)$ 矩阵. 如果 $m<n$，证明是类似的. □

练习 设 $A\in M_{m,n}$ 是给定的矩阵. 证明 A^*、A^T 和 $\bar{A}$ 的奇异值与 A 的相同. 如果 $U\in M_m$ 和 $V\in M_n$ 是酉矩阵，证明 UAV 的奇异值与 A 的相同. 如果 $c\in\mathbf{C}$，证明 cA 的奇异值是 A 的奇异值的 $|c|$ 倍.

作为定理(7.3.7)的一个直接应用，我们有关于任意矩阵的奇异值的一些扰动结果，这些结果是从 Hermite 矩阵的相应结果得来的. 它们说明每个矩阵关于奇异值的计算是优态的，可以把它们同(6.3.2)、(6.3.4)以及其中关于条件数的讨论作一比较. 关于如何把这些结果推广到任意酉不变范数，可参看(7.4.51). 418

7.3.8 推论 设 $A, B\in M_{m,n}$，$E\equiv B-A$，且设 $q=\min\{m, n\}$. 如果 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_q$ 是 A 的奇异值，且 $\tau_1\geqslant\tau_2\geqslant\cdots\geqslant\tau_q$ 是 B 的奇异值则

(a) $|\sigma_i-\tau_i|\leqslant\|E\|_2$ 对所有 $i=1, 2, \cdots, q$ 成立；

(b) $\left[\sum_{i=1}^{q}(\sigma_i-\tau_i)^2\right]^{1/2}\leqslant\|E\|_2$.

证明： 这两个结果类似于 Weyl 不等式[(4.3.1)；也可参看(6.3.5)前面的练习]以及关于 Hermite 矩阵的 Hoffmam-Wielandt 定理(6.3.8). 它们可从所述结果及(7.3.7)直接得出. □

练习 给出(7.3.8)的详细证明. 关于(a)见(5.6)节习题 36.

奇异值也有交错性质；它是由 Hermite 矩阵的特征值交错性质得来的.

7.3.9 定理 设 $A\in M_{m,n}$ 是给定的矩阵，又设 $\hat{A}$ 是划去 A 的任意一列后得到的矩阵. 设 $\{\sigma_i\}$ 表示 A 的奇异值，$\{\hat{\sigma}_i\}$ 表示 $\hat{A}$ 的奇异值，且都按递减顺序排列.

(a) 如果 $m\geqslant n$，则

$$\sigma_1\geqslant\hat{\sigma}_1\geqslant\sigma_2\geqslant\hat{\sigma}_2\geqslant\cdots\geqslant\hat{\sigma}_{n-1}\geqslant\sigma_n\geqslant 0.$$

(b) 如果 $m<n$，则

$$\sigma_1\geqslant\hat{\sigma}_1\geqslant\sigma_2\geqslant\hat{\sigma}_2\geqslant\cdots\geqslant\sigma_m\geqslant\hat{\sigma}_m\geqslant 0.$$

如果划去 A 的一行，而不是一列，则相应的不等式可通过交换(a)和(b)中的 m 和 n 来得到.

证明：A 的奇异值的平方是 Hermiter 矩阵 $A^*A\in M_n$ 的特征值，而 $\hat{A}$ 的奇异值的平方是 $\hat{A}^*\hat{A}\in M_{n-1}$ 的特征值；如果划去 A 的一列，$\hat{A}^*\hat{A}$ 是 A^*A 的主子矩阵. 交错不等式组可从包含原理(1.3.15)直接得出. 如果划去 A 的一行而不是一列，则相应地考虑 AA^* 和 $\hat{A}\hat{A}^*$. □

119

作为 Hermite 矩阵的特征值性质与奇异值性质间的最后一个类似，有与 Courant-Fischer 定理(4.2.11)类似的定理.

7.3.10 定理 设 $A\in M_{m,n}$，$q=\min\{m, n\}$，设 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_q$ 是 A 的有序奇异值，又设 k 是适合 $1\leqslant k\leqslant q$ 的某个整数. 则

$$\min_{w_1,\cdots,w_{k-1}\in\mathbf{C}^n}\ \max_{\substack{x\neq 0,x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{k-1}}}\frac{\|Ax\|_2}{\|x\|_2}=\sigma_k,$$

且

$$\max_{w_1,\cdots,w_{n-k}\in\mathbf{C}^n}\ \min_{\substack{x\neq 0,x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{n-k}}}\frac{\|Ax\|_2}{\|x\|_2}=\sigma_k.$$

证明：这两个公式可直接从(4.2.12)和(4.2.13)推出，因为 $\sigma_k^2(A)$是 A^*A 的特征值. 如果 $\lambda_1\leqslant\lambda_2\leqslant\cdots\lambda_k$ 是 Hermite 矩阵 A^*A 的有序特征值，则 $\sigma_k^2(A)=\lambda_{n-k+1}(A^*A)$，而(4.2.12)说明

$$\sigma_k^2(A)=\lambda_{n-k+1}(A^*A)=\min_{w_1,\cdots,w_{k-1}\in\mathbf{C}^n}\ \max_{\substack{x\neq 0,x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{k-1}}}\frac{x^*A^*Ax}{x^*x}$$

$$=\min_{w_1,\cdots,w_{k-1}\in\mathbf{C}^n}\ \max_{\substack{x\neq 0,x\in\mathbf{C}^n\\ x\perp w_1,\cdots,w_{k-1}}}\left(\frac{\|Ax\|_2}{\|x\|_2}\right)^2.$$

用同样的方法可以证明第二个恒等式. □

习题

1. 设 $P\in M_n$ 是半正定矩阵. 证明 P 可以写成 P^2 的多项式，因此，如果某个矩阵 U 与 P^2 可交换，则它一定也与 P 可交换. 用这个事实证明，如果 $A\in M_n$ 是正规矩阵，则它的极因子 P 和 U 可交换.

2. 证明，任意 $A\in M_n$ 可写成 $A=Pe^{iH}$，其中 P，$H\in M_n$，P 是半正定矩阵，而 H 是 Hermite 矩阵. 证明 H 可以取为正定矩阵. P 和 H 可由 A 确定到什么程度？**提示**：如果 $U\in M_n$ 是酉矩阵，且 $U=V\Lambda V^*$ 是 U 的酉对角化，则 $\Lambda=e^{iD}$，其中 D 是具有实主对角元的对角矩阵. e^{iVDV^*} 是什么？

120

3. 证明，$A\in M_n$ 有零奇异值，当且仅当它有零特征值.

4. 设 $A\in M_{m,n}$ 且 $q=\min\{m, n\}$. 证明 A 的最大奇异值等于 A 的谱范数. 证明 A 的 Frobenius 范数适合恒等式

$$\|A\|_2=\left(\sum_{i=1}^{q}\sigma_i^2\right)^{1/2}$$

证明 $\sigma_1\leqslant\|A\|_2\leqslant\sqrt{n}\sigma_1$，并且确定使等式成立的条件. 证明对所有 $A\in M_n$ 有

$$|\!|\!|A|\!|\!|_2\leqslant\|A\|_2\leqslant\sqrt{n}\,|\!|\!|A|\!|\!|_2. \tag{7.3.11}$$

考虑 I 和 $\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$，说明 $\|A\|_2$ 可以达到这两个界.

5. 如果 $k\leqslant\min\{m, n\}$，且在 A 的奇异值分解(7.3.5)中，v_k 是 V 的第 k 列，w_k 是 W 的第 k 列，证明

$$A^* v_k = \sigma_k w_k \quad 和 \quad Aw_k = \sigma_k v_k,$$

其中 σ_k 是 A 的第 k 个奇异值. 特别是，$v_k^* Aw_k=\sigma_k$.

6. 如果给出一个大矩阵 A，如何用数值方法去计算 A 的秩呢？注意到 A 的秩等于 A 的非零奇异值的个数，于是，一种从数值上计算 A 的秩的方法是：求奇异值分解，然后取 A 的秩为大于某个限定值的奇异值的个数. 如果 A 的最小非零奇异值与最大非零奇异值之比不接近于 0，你为什么可以指望用数值方法确定 A 的秩会更容易和更精确？

7. 设 $A\in M_{m,n}$ 有奇异值分解 $A=V\Sigma W^*$，且定义 $A^\dagger=W\Sigma^\dagger V^*$，其中，$\Sigma^\dagger$ 是 Σ 的转置，且 Σ 中 A 的各正奇异值用它们的倒数来代替. 证明，

(a) $AA^\dagger$ 和 $A^\dagger A$ 是 Hermite 矩阵；

(b) $AA^\dagger A=A$；

(c) $A^\dagger AA^\dagger=A^\dagger$.

证明，如果 A 是非奇异方阵，则 $A^\dagger=A^{-1}$. 矩阵 $A^\dagger$ 称为 A 的 Moore-Penrose 广义逆. 对于任意矩阵 A，甚至于奇异方阵 A 和非方阵 A，它们都有广义逆. 再证明，按上述要求(a)～(c)，$A^\dagger$ 是唯一确定的.

8. 线性方程组 $Ax=b$ 的最小二乘解是在使 $\|Ax-b\|_2$ 为极小的所有向量 x 当中，$\|x\|_2$ 达到极小值的向量 x. 证明 $x=A^\dagger b$ 是 $Ax=b$ 的最小二乘解.

421

9. 证明 $A^\dagger=\lim\limits_{t\to 0}A^*(AA^*+tI)^{-1}$，其中 $A^\dagger$ 是习题 7 中所定义的矩阵.

10. 不直接利用特征向量和特征值也可以导出奇异值分解(7.3.5). 回顾一下利用 Rayleigh-Ritz 原理对角化一个 Hermite 矩阵的证明，从谱范数的变分特征可以直接构造出(左和右)奇异向量和奇异值. 考虑 $A\in M_n$ 及变分特征(∗∗) $\|A\|_2=\max\{\|Ax\|_2: \|x\|_2=1\}$.
(a)设 $n\geqslant 2$，又设 $B\in M_n$ 有特殊形式

$$B=\begin{bmatrix}\sigma_1 & w^*\\ 0 & X\end{bmatrix}$$

其中 $\sigma_1=\|B\|_2$，$w\in\mathbf{C}^{n-1}$ 而 $X\in M_{n-1}$. 证明 $w=0$. **提示：** 如果 $\sigma_1>0$，考虑 $\zeta\equiv\begin{bmatrix}\sigma_1\\ w\end{bmatrix}/(\sigma_1^2+w^*w)^{1/2}$，证明 $\|B\zeta\|_2^2\geqslant\sigma_1^2+w^*w$，然后利用(∗∗). (b)设 $A\in M_n$ 且 $\sigma_1\equiv\|A\|_2$. 然后利用(∗∗)证明，存在某个单位向量 x_1 使得 $\|Ax_1\|_2=\sigma_1$. 设 $y_1\equiv\sigma_1^{-1}Ax_1$. (c)设 W_1，$V_1\in M_n$ 是酉矩阵，它们的第 1 列分别是 x_1 和 y_1. 证明 $V_1^*AW_1$ 有谱范数 σ_1 且有(a)中的矩阵形式. 由此得出 $V_1^*AW_1=\begin{bmatrix}\sigma_1 & 0\\ 0 & X\end{bmatrix}$. (d)通过引入非对角零元组成的列和行压缩 A，用公式表示这个归纳过程，左乘和右乘相应的酉矩阵便可得到 A 的奇异值分解. (e)如果 $A\in M_{m,n}$ 不是方阵，情况又如何？

11. 设 $A=V\Sigma W^*$ 是矩阵 $A\in M_{m,n}$ 的奇异值分解，假定 A 有秩 k，且设 $q=\min\{m, n\}$. 证明 W 的最后 $n-k$ 列构成 A 的零空间的标准正交基，而 V 的前 k 列构成 A 的值域的标准正

交基.

12. 设 $A\in M_{m,n}$，$B\in M_{p,n}$. 证明 A 和 B 的零空间的交的标准正交基由 W 的后若干列(有多少?)给出，其中 $V\Sigma W^*$ 是分块矩阵 $\begin{bmatrix}A\\B\end{bmatrix}\in M_{m+p,n}$ 的奇异值分解. **提示：**对 $x\in\mathbf{C}^n$，什么时候 $\begin{bmatrix}A\\B\end{bmatrix}x=0$? 你如何求列数相同的 k 个矩阵 A_1，A_2，$\cdots$，A_k 的各个零空间的交的标准正交基.

13. 证明极分解(7.3.2)与奇异值分解在容易相互导出的意义下是等价的. **提示：**把谱定理应用于 P.

[422]

14. 设 $A\in M_n$. 证明，A 可对角化，当且仅当存在正定 Hermite 矩阵 P 使得 $P^{-1}AP$ 是正规矩阵. **提示：**如果 $A=SAS^{-1}$，应用极分解(7.3.3)于 S.

15. 试用奇异值分解(7.3.5)(特别是关于分解的唯一性的论述)以及推论(7.3.6)证明，Takagi 表示(4.4.4)对复对称矩阵成立. **提示：**若 $A=A^T\in M_n$ 有各不相同的奇异值且 $A=V\Sigma W^*$，则 $A=A^T=\bar{W}\Sigma V^T$. 另一方面，存在一个对角酉矩阵 $D=\mathrm{diag}(e^{i\theta_1},\cdots,e^{i\theta_n})$ 使得 $\bar{W}=VD$，故 $A=V\Sigma W^*=V\Sigma(\overline{VD})^*=V\Sigma DV^T=(VD^{1/2})\Sigma(VD^{1/2})^T\equiv U\Sigma U^T$. 对于一般情形，把(7.3.6)和选择原理(2.1.8)用于扰动，然后取极限.

16. 设 A，$B\in M_{m,n}$，$q=\min\{m, n\}$，设 A 的有序奇异值是 $\sigma_1(A)\geqslant\cdots\geqslant\sigma_q(A)\geqslant 0$. 类似地，设 B 和 $A+B$ 也有有序奇异值. 设 $\tilde{A}$、$\tilde{B}$、$\tilde{A}+\tilde{B}\in M_{m+n}$ 是如(7.3.7a)中定义的 Hermite 矩阵. 证明，对 $k=1, 2, \cdots, q$ 有 $\sigma_k(A)=\lambda_{m+n-k+1}(\tilde{A})$；且对 B 和$(A+B)$也有类似的结果. 注意：诸奇异值是按递减顺序排列的，而 Hermite 矩阵 $\tilde{A}$ 的诸特征值是按递增顺序排列的. 试用这个等式和 Weyl 定理(4.3.7)证明

$$\sigma_{i+j-1}(A+B)\leqslant\sigma_i(A)+\sigma_j(B),\quad 1\leqslant i,j\leqslant q \quad 和 \quad i+j\leqslant q+1.$$

特别地，$\sigma_1(A+B)\leqslant\sigma_1(A)+\sigma_1(B)$，(为什么这是意料之中的?)且 $\sigma_q(A+B)\leqslant\min\{\sigma_q(A)+\sigma_1(B), \sigma_1(A)+\sigma_q(B)\}$.

17. 考察 $A=\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$ 和 $B=\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$，证明不等式 $\sigma_i(A+B)\leqslant\sigma_i(A)+\sigma_i(B)$ 不是对所有 $i=1, 2$ 都成立，其中$\{\sigma_i(A)\}$和$\{\sigma_i(B)\}$分别是 A 和 B 的奇异值，且都按递降顺序排列.

18. 设 A，$B\in M_{m,n}$ 是给定的，$q=\min\{m, n\}$，设 A 的有序奇异值是 $\sigma_1(A)\geqslant\cdots\geqslant\sigma_q(A)\geqslant 0$，且 B 和 $AB^*\in M_m$ 的奇异值也有类似的递减顺序. 证明

$$\sigma_{i+j-1}(AB^*)\leqslant\sigma_i(A)\sigma_j(B),\quad 1\leqslant i,j\leqslant q,\quad i+j\leqslant q+1.$$

这些不等式可以看作类似于习题 16 中的加法不等式的乘法不等式，当 $m=n$ 时，它们也可以看作谱范数的次乘性质的推广. 为什么? **提示：**设 $AB^*=WQ$ 是 AB^* 的左极分解，其中，$W\in M_m$ 是酉矩阵，$Q\in M_m$ 是半正定矩阵. 对任一个 $x\in\mathbf{C}^m$，有 $(x^*Qx)^2=(x^*W^*AB^*x)^2=|(A^*Wx)^*(B^*x)|^2\leqslant\|A^*Wx\|_2^2\|B^*x\|_2^2=[(Wx)^*AA^*(Wx)](x^*BB^*x)$. 设 z_1，$\cdots$，z_{i-1} 是 AA^* 的标准正交特征向量，它们对应于 AA^* 的 $i-1$ 个最大特征值 $\sigma_1^2(A)$，$\cdots$，$\sigma_{i-1}^2(A)$，设 y_1，$\cdots$，y_{j-1} 是 BB^* 的标准正交特征向量，它们对应于 BB^* 的 $j-1$ 个最大特征值 $\sigma_1^2(B)$，$\cdots$，$\sigma_{j-1}^2(B)$，又设 $x_1=W^*z_1$，$x_2=W^*z_2$，$\cdots$，$x_{i-1}=W^*z_{i-1}$，$x_i=y_1$，$x_{i+1}=y_2$，$\cdots$，$x_{i+j-2}=$

[423]

y_{j-1}. 若对 $k=1,2,\cdots,i+j-2$，x 都正交于 x_k，则 $(Wx)^*AA^*(Wx)\leqslant\sigma_i^2(A)\|x\|_2^2$ 和 $x^*BB^*x\leqslant\sigma_j^2(B)\|x\|_2^2$，因此，在这些限制下，我们有 $(x^*Qx)^2\leqslant\sigma_i^2(A)\sigma_j^2(B)\|x\|_2^4$. 于是引用 Courant-Fischer 定理(4.2.11)便得出

$$\sigma_{i+j-1}^2(AB^*) = (\lambda_{n-i-j+2}([(AB^*)^*(AB^*)]^{1/2}))^2 \leqslant \sigma_i^2(A)\sigma_j^2(B).$$

19. 虽然当 $A,B\in M_n$ 时，AB 和 BA 的特征值总是相同的，不过例子 $\begin{bmatrix}0&1\\0&0\end{bmatrix}$ 和 $\begin{bmatrix}0&0\\0&1\end{bmatrix}$ 却说明，AB 和 BA 的奇异值不一定相同. 但是可以证明 AB 和 B^*A^* 的奇异值总是相同的.

20. 设 X 是 n 维随机向量，它的诸分量有零平均值和有限方差. 设 $\Sigma\equiv\mathrm{Cov}(X)=E(XX^*)$[见(4.5.3*)]，假定 Σ 是非奇异的，设 $P=\Sigma^{1/2}$，且 $A,B\in M_n$ 是给定的，随机向量 AX 和 BX 有相同的(零)平均值向量，但没有理由指望它们有相同的协方差矩阵. 证明，$\mathrm{Cov}(AX)=\mathrm{Cov}(BX)$ 当且仅当 $A=B(PUP^{-1})$ 对某个酉矩阵 $U\in M_n$ 成立. **提示**：若 $A\Sigma A^*=B\Sigma B^*$，则 $(AP)(AP)^*=(BP)(BP)^*$. 如果 RW 是 BP 的一个极分解，证明，对某个酉矩阵 $W,V\in M_n$，RV 是 AP 的一个极分解. R 是什么? 由此得出 $A=B(PW^*VP^{-1})=B(PUP^{-1})$. U 可以确定到什么程度? 若 $\Sigma=I$ 又如何? 若 $B=I$ 又如何?

21. 考虑由

$$A_\varepsilon=\begin{bmatrix}0&1&&0\\ \vdots&\ddots&\ddots&\\ 0&&\ddots&1\\ \varepsilon&0&\cdots&0\end{bmatrix},\quad \varepsilon>0$$

给出的矩阵 $A_\varepsilon\in M_n$. 证明 A_ε 的特征多项式是 $t^n-\varepsilon$. **提示**：用沿其第 1 列的 Laplace 代数余子式展开计算 $\det(tI-A_\varepsilon)$. 证明 A_ε 的各特征值是 $\sqrt[n]{\varepsilon}$ 的 n 个选择. 证明 A_ε 的各奇异值是 $1(n-1$ 重)和 ε. 现在设 $n=10$，$\varepsilon=10^{-10}$，注意到扰动 $A_0\to A_\varepsilon$ 引起 A_0 的各特征值一个 0.1 扰动，但只引起 A_0 的任一奇异值一个 10^{-10} 的扰动. A_ε 的谱条件数是什么? 这是关于定理(7.3.7)下面的论断的一个例子，该论断是说，每个矩阵关于奇异值的计算是良态的，而一个给定矩阵关于特征值的计算可能是病态的. [424]

22. 设 $A=[a_{ij}]\in M_n$ 是给定的. 证明，若 A 有"小"行或列，则 A 一定有"小"奇异值. 说得更明确些，设 $A=[r_1r_2\cdots r_n]^T$，其中 $r_i\in\mathbf{C}^n$，且 r_i^T 是 A 的第 i 行. 用递增顺序排列各行的 Euclid 范数 $\{\|r_i\|_2:\ i=1,\cdots,n\}$，并且用 $R_1\leqslant R_2\leqslant\cdots\leqslant R_n$ 表示所得到的有序值. 证明

$$\sum_{i=1}^k\sigma_{n-i+1}^2\leqslant\sum_{i=1}^kR_i^2,\quad k=1,2,\cdots,n;$$

一个类似的上界可用列范数表示. 注意诸奇异值是取顺序 $\sigma_n\leqslant\sigma_{n-1}\leqslant\cdots\leqslant\sigma_1$. **提示**：平方奇异值是 Hermite 矩阵 AA^* 的特征值. AA^* 的各主对角元是什么? 用优化和定理(4.3.26). 关于列和不等式考虑 A^*A. 与(4.3)节习题 19 比较.

23. 有一个与奇异值分解类似的自然分解，其中的酉因子用复正交因子来代替. 但是，与奇异值分解不同的是，这个分解不是总能实现的：从(2.3)节习题 7 可以想到，类似于 Schur 酉上三角分解的正交分解也不是总能实现的. 如果 $A\in M_{m,n}$ 可以写成 $A=P\Lambda Q^T$ 的形式，其中

$P\in M_m$ 和 $Q\in M_n$ 是复正交矩阵，而 $\Lambda=[\lambda_{ij}]\in M_{m,n}$ 是在 $i\neq j$ 时有 $\lambda_{ij}=0$ 的意义下的"对角"矩阵，证明 $AA^T\in M_m$ 是可对角化的且 rank $A=$ rank AA^T. 这两个条件也足以确保所述分解 $A=P\Lambda Q^T$ 的存在性. 如果 A 是实矩阵，这指的是什么？给出一个 $A\in M_2$ 的例子，说明它不能写成 $A=P\Lambda Q^T$，其中 P，$Q\in M_2$ 是复正交矩阵，而 $\Lambda\in M_2$ 是对角矩阵.

24. 说明奇异值分解为什么可以看作正规矩阵的谱定理的推广.

25. 关于正规矩阵交换族同时酉对角化的定理(2.5.5)对于奇异值分解有一个类似的结果. 设 $\mathscr{F}=\{A_i: i\in\mathscr{I}\}\subset M_{m,n}$，假定存在酉矩阵 $V\in M_m$ 和 $W\in M_n$ 使得每个 V^*A_iW 是在习题 23 的意义下(即在 $i\neq j$ 时其 i，j 元为 0)的"对角"矩阵. 证明，(a)每个 $A_i^*A_j\in M_n$ 是正规矩阵，且 $\varsigma=\{A_iA_j^*: i, j\in\mathscr{I}\}\subset M_m$ 是一个交换族. (b) $A_iA_j^*A_k=A_kA_j^*A_i$ 对每个 i，j，$k\in\mathscr{I}$ 成立. 这每一个必要条件也是族 $\mathscr{F}$ 有奇异值分解形式的同时分解的充分条件.

[425]

26. 求两个给定矩阵 A，$B\in M_{m,n}$ 的奇异值分解形式的同时分解是前一个习题的一个有意思的特殊情形. 证明，存在酉矩阵 $V\in M_m$，$W\in M_n$ 使得 $A=V\Sigma W^*$ 和 $B=V\Lambda W^*$ (其中 Σ，$\Lambda\in M_{m,n}$ 是"对角"矩阵)当且仅当 AB^* 和 B^*A 都是正规矩阵. **提示**：为了证明该条件是充分的，可证明只要考虑 $A=\Sigma$ 是非负矩阵又是"对角"矩阵的情形. 若把 Σ 的相同对角元排在一块，证明，若 ΣB^* 和 $B^*\Sigma$ 是正规矩阵，则 B 是一个分块对角矩阵，其中可能除了一个子块以外(若 A 是奇异矩阵)所有子块都是正规矩阵. 对每一个子块，或者用关于正规矩阵的谱定理或者用奇异值分解便可得到结论.

27. 如果我们希望有酉矩阵 $V\in M_m$ 和 $W\in M_n$ 使得族 $\mathscr{F}=\{A_i: i\in\mathscr{I}\}\subset M_{m,n}$ 中的每一个成员都可以写成 $A_i=V\Sigma_iW^*$，其中每个 Σ_i 都是"对角"矩阵，证明，对每个 i，$j\in\mathscr{I}$，$A_iA_j^*\in M_m$ 和 $A_i^*A_j\in M_n$ 都是正规矩阵的条件是必要的，但是当该族有三个或更多的矩阵时，它不是充分的. **提示**：考虑

$$\mathscr{F}=\left\{\begin{bmatrix}1&0\\0&i\end{bmatrix},\begin{bmatrix}0&1\\1&0\end{bmatrix},\begin{bmatrix}0&1\\-1&0\end{bmatrix}\right\}.$$

当该族的矩阵多于两个时，说明关于两个矩阵情形的证明中的哪个部分行不通.

进一步阅读与注释 Sylvester 于 1889 年证明了关于实方阵的奇异值分解. 关于一般的 $m\times n$ 复矩阵的奇异值分解的最早证明似乎是在下文中：C. Eckart and G. Young, "A Principal Axis Transformation for Non-Hermitian Matrices," *Bull. Amer. Math. Soc.* 45 (1939), 118-121. Eckart 和 Young 的文章也包含了如下结果：两个矩阵 A，$B\in M_{m,n}$ 有奇异值分解形式的同时分解(其中的相应"对角"因子都是实矩阵)当且仅当 AB^* 和 B^*A 都是 Hermite 矩阵. 关于矩阵族有奇异值分解形式的同时分解的诸多结果的一个综述以及更多的参考资料可参看 P. M. Gibson, "Simultaneous Diagonalization of Rectangular Complex Matrices," *Linear Algebra Appl.* 9(1974), 45-53.

[426]

7.4 奇异值分解的例子和应用

极分解和奇异值分解有许多应用. 一些应用在习题中给出，而一些应用在下述例子中讨论.

7.4.1 例 如果 $A\in M_n$ 是给定的可逆矩阵，则(关于任一范数)与 A 充分接近的所有矩阵也是可逆的. 在某些统计模型问题中，需要求一个在最小二乘意义下与 A"最接近的奇异矩阵"；也就是希望求矩阵 B 使得 $A+B$ 是奇异矩阵，而 $\|B\|_2$ 越小越好.

设$\|\cdot\|$是任一矩阵范数. 考虑 $A+B=A(I+A^{-1}B)$，假定它是奇异矩阵. 假如$\|A^{-1}B\|<1$，由(5.6.16)可知，$I+A^{-1}B$，因而 $A+B$ 是可逆的. 于是 $1\leqslant\|A^{-1}B\|\leqslant\|A^{-1}\|\|B\|$，因此，如果 $A+B$ 是奇异矩阵而 A 是可逆矩阵，我们就一定有$\|B\|\geqslant 1/\|A^{-1}\|$. 如果选$\|\cdot\|$为谱范数，又如果 $A=V\Sigma W^*$ 是 A 的奇异值分解，则$\|A^{-1}\|_2=\|W\Sigma^{-1}V^*\|_2=\|\Sigma^{-1}\|_2=1/\sigma_n$，其中 σ_n 是 A 的最小奇异值. 于是，使得 $A+B$ 是奇异矩阵的任何 B 必须适合$\|B\|_2\geqslant\sigma_n(A)$. 但是，如果选择 B 为矩阵 $B=VEW^*$，其中 $E=\mathrm{diag}(0,0,\cdots,0,-\sigma_n)$，则$\|B\|_2=\|E\|_2=\sigma_n=\|E\|_2=\|B\|_2$，且 $A+B$ 就是奇异矩阵(有秩 $n-1$).

更一般地，如果关于 Frobenius 范数需要求一个与某个奇异或非奇异矩阵 A"最接近的秩 k 矩阵"，可以这样选择 $A+B$，其中 $B=VEW^*$ 如前，但是 $E=\mathrm{diag}(0,\cdots,0,-\sigma_{k+1},\cdots,-\sigma_n)$. 关于这个结果从 Frobenius 范数到所有酉不变范数的推广，可参看本节末习题 1 的有关证明以及例(7.4.52).

$k=1$ 的情形经常出现在一些应用中，这是值得特别提出来的. 用秩 1 矩阵 $X\in M_n$ 对某个矩阵 $A=V\Sigma W^*\in M_n$ 的最佳最小二乘逼近是 $X=A+B=V(\Sigma+E)W^*=V\mathrm{diag}(\sigma_1,0,\cdots,0)W^*=\sigma_1 vw^*$，其中，$\sigma_1$ 是 A 的最大奇异值，而 v 和 w 分别是 A 的奇异值分解中的酉矩阵 V 和 W 的第 1 列. 关于 v 和 w 的一个有用的论断是，v 和 w 是一对 Hermite 特征值—特征向量问题

$$AA^*v=\sigma_1^2v,\quad A^*Aw=\sigma_1^2w$$

的单位向量解，其中 σ_1^2 是半正定矩阵 A^*A(和 AA^*)的最大特征值. 这个论断当然不唯一确定 v 和 w；一个困难是相应于 σ_1^2 的特征空间不一定是一维的. 但是，如果 σ_1^2 是 A^*A(因而是 AA^*)的单特征值，则向量 v 和 w 可确定到相差模为 1 的纯量因子，因此，v 和 w 一定是奇异值分解 $A=V\Sigma W^*$ 中酉矩阵 V 和 W 的相应第 1 列的纯量倍. 在这个意义下，关于单位特征向量 v 和 w 的确定选择，A 的最佳秩 1 逼近必定是形式 $e^{i\theta}\sigma_1 vw^*$，其中 $\theta\in\mathbf{R}$ 是某个实数. 我们必须选取纯量因子 $e^{i\theta}$ 使得 $\|A-e^{i\theta}\sigma_1vw^*\|_2^2=\|A\|_2^2-2\sigma_1\mathrm{Re}[\mathrm{tr}\, e^{-i\theta}A(vw^*)^*]+\sigma_1^2\|v\|_2^2\|w\|_2^2$为极小，问题等价于使 $\mathrm{Re}[\mathrm{tr}\, e^{-i\theta}A(vw^*)^*]=\mathrm{Re}[e^{-i\theta}v^*Aw]$为极大. 但是，$Aw=V\Sigma W^*w=e^{i\phi}\sigma_1 v$ 对某个 $\phi\in\mathbf{R}$ 成立[见(7.3)节习题 5]，因而 $|v^*Aw|=\sigma_1>0$. 因此，最佳纯量因子是 $e^{i\theta}=v^*Aw/|v^*Aw|=v^*Aw/\sigma_1$，而对 A 的最佳秩 1 逼近是

$$e^{i\theta}\sigma_1vw^*=(v^*Aw)vw^*.$$

427

这说明，如果 A^*A 的最大特征值是单的，则对 A 的最佳秩 1 最小二乘逼近可以毫不费力地通过两个 Hermite 特征值问题的解构造出来. 例如，使 AA^T 为正矩阵，或更一般地为不可约矩阵的任一非负矩阵 $A\in M_n(\mathbf{R})$就适合 A^*A 的最大单特征值的条件[见(8.4)节习题 17].

7.4.2 例 在定理(5.7.17)中已经证明，M_n 上的向量范数 $G(\cdot)$能使条件

$$G(A_1)G(A_2)\cdots G(A_k)\geqslant\rho(A_1\cdots A_k)$$

对所有 A_1, A_2, …, $A_k \in M_n$ 和所有 $k=1, 2, \cdots$ 成立，当且仅当 $G(\cdot)$ 在 $\mathbf{C}^n$ 上有相容向量范数. 在这个证明中，决定性的一步是证明，如果 $G(\cdot)$ 满足这个关于谱半径的不等式，则有某个有限常数 $c>0$，使得 $G(A_1)G(A_2)\cdots G(A_k) \geqslant c\|A_1A_2\cdots A_k\|_2$，而证明它的关键是乘积 $A_1A_2\cdots A_k$ 的奇异值分解. 其细节在引理(5.7.16)中.

7.4.3 例 假定想解线性方程组 $Ax=b$，其中，$A\in M_{m,n}$ 和 $b\in \mathbf{C}^m$ 是已知的，且有秩 k. 如果 $A=V\Sigma W^*$ 是 A 的奇异值分解，则 $V\Sigma W^* x=b$，或

$$\Sigma(W^* x) = V^* b \tag{7.4.4}$$

如果 $m>k$，则 Σ 的后 $m-k$ 行是 0，因此，如果在这种情形下有解，则必须(同时也只须)使 [428] $V^* b$ 的后 $m-k$ 个元为零. 于是，方程组在 $m>k$ 时有解，当且仅当 b 与 A 的后 $m-k$ 个左奇异向量正交. 如果 b 满足这个相容条件，又如果 $V=[v_1\cdots v_m]$ 和 $W=[w_1\cdots w_n]$，则(7.4.4)说明

$$(W^* x)^* = \left[\frac{b^* v_1}{\sigma_1},\cdots,\frac{b^* v_k}{\sigma_k},0,\cdots,0\right]^*,$$

因而向量

$$x = \sum_{i=1}^{k} \frac{v_i^* b}{\sigma_i} w_i \tag{7.4.5}$$

是解. 因为对所有 $j>k$，$Aw_j=V(\Sigma W^* w_j)=0$，所以，A 的后 $n-k$ 个右奇异向量(如果有的话)的任意线性组合都在 A 的零空间中，因而对任意 c_{k+1}, …, $c_n\in\mathbf{C}$，向量

$$x = \sum_{i=1}^{k} \frac{v_i^* b}{\sigma_i} w_i + \sum_{i=k+1}^{n} c_i w_i$$

都是 $Ax=b$ 的解；当然，如果 $n=k$，就不会出现这后一个和式. 因为向量组 $\{w_i\}$ 是标准正交组，所以当所有 $c_i=0$ 时，就得到具有极小 l_2 范数的解. 值得指出的是，A 的后 $m-k$ 个左奇异向量张成 AA^* 的零空间，它与 A^* 的零空间相同，因此，要求 b 与 A 的后 $m-k$ 个奇异向量正交与要求 b 与 $A^* x=0$ 的每个解正交是一回事.

练习 如果 $V^* b$ 的后 $m-k$ 个元不全为零，则方程组 $Ax=b$ 是不相容的，因而根本没有解，但是，为了某些目的. 只需要有"最小二乘"解就可以了. 这个解是使 $\|Ax-b\|_2$ 达到极小的，具有极小 l_2 范数的向量 x. 证明(7.4.5)给出这样一个最小二乘解.

7.4.6 例 用酉矩阵的纯量倍对某个矩阵 $A\in M_n$ 的最佳最小二乘逼近是什么？我们知道，M_n 上的 l_2 范数是由内积 $[A, B]=\mathrm{tr}\,AB^*$ 诱导的，还知道，如果 U 是酉矩阵，则

$$\|U\|_2^2 = [U,U] = \mathrm{tr}\,UU^* = \mathrm{tr}\,I = n$$

对任意 $c\in\mathbf{C}$ 和任意酉矩阵 $U\in M_n$，有

$$\|A-cU\|_2^2 = [A-cU, A-cU] = \|A\|_2^2 - 2\mathrm{Re}(\bar{c}[A,U]) + n|c|^2,$$

当 $c=[A, U]/n$ 时它达到极小，因而

[429]

$$\|A-cU\|_2^2 \geqslant \|A\|_2^2 - \frac{1}{n}|[A,U]|^2.$$

如果定义

$$u(A)=\max_{\text{酉矩阵}U\in M_n}|[A,U]|. \tag{7.4.7}$$

就得到一个与数值半径 $r(A)$ 类似的量. 对于 $r(A)$，内积的极大值不是针对酉矩阵来取的，而是取遍所有 Frobenius 范数为 1 的秩 1Hermite 矩阵. 但是，与数值半径不同的是，函数 $u(A)$ 是 M_n 上的矩阵范数[见习题 5 和例(7.4.54)].

要确定 $u(A)$ 的值与欲求的酉矩阵并不困难. 设 A 的奇异值分解是 $A=V\Sigma W^*$. 则

$$\begin{aligned}u(A)&=\max_{\text{酉矩阵}U}|[A,U]|=\max_{\text{酉矩阵}U}|[V\Sigma W^*,U]|\\&=\max_{\text{酉矩阵}U}|\operatorname{tr}V\Sigma W^*U^*|=\max_{\text{酉矩阵}U}|\operatorname{tr}\Sigma(W^*U^*V)|\\&=\max_{\text{酉矩阵}U}|\operatorname{tr}\Sigma U|=\max_{\text{酉矩阵}U=[u_{ij}]}\left|\sum_{i=1}^{n}\sigma_i u_{ii}\right|\\&\leqslant\max_{\text{酉矩阵}U=[u_{ij}]}\sum_{i=1}^{n}\sigma_i|u_{ii}|\leqslant\sum_{i=1}^{n}\sigma_i.\end{aligned}$$

但是，如果 $A=PU$ 是 A 的极形式，则

$$[A,U]=\operatorname{tr}PUU^*=\operatorname{tr}P=\sum_{i=1}^{n}\sigma_i.$$

因此，所给出的 $u(A)$ 的上界是可以达到的，$u(A)=\sigma_1(A)+\cdots+\sigma_n(A)$. 又如果 $A=PU$ 是 A 的极形式，且 σ_1，…，σ_n 是它的奇异值，则用一个酉矩阵的倍数对 A 的最佳最小二乘逼近可由

$$\frac{1}{n}(\sigma_1+\cdots+\sigma_n)U$$

给出. 如果给定奇异值分解 $A=V\Sigma W^*$，则 $U=VW^*$，该逼近的误差是

$$\left\|A-\frac{u(A)}{n}U\right\|_2^2=\|A\|_2^2-\frac{1}{n}|[A,U]|^2=\sum_{i=1}^{n}\sigma_i^2-\frac{1}{n}\Big(\sum_{i=1}^{n}\sigma_i\Big)^2,$$

只有当 Cauchy-Schwarz 不等式

$$\Big(\sum_{i=1}^{n}\sigma_i 1\Big)^2\leqslant\Big(\sum_{i=1}^{n}1^2\Big)\Big(\sum_{i=1}^{n}\sigma_i^2\Big)$$

是等式时误差才为 0. 因此，只有当 A 的所有奇异值都相同时，A 才可以用一个酉矩阵的倍数来完全逼近.

430

7.4.8 例 假定 $A,B\in M_{m,n}$ 是给定的矩阵，而我们想知道，是否可以通过"旋转"B 来得到 A；即 $A=UB$ 对某个酉矩阵 $U\in M_m$ 成立吗？更一般地，如果考虑已知矩阵 B 的所有可能的"旋转"UB，在最小二乘意义下，可以怎样充分地逼近 A？这个问题在因子分析中称为求 B 的一个"强行(procrustean)变换"问题.

要做的计算与上例中的计算是类似的；我们试图选择 U 使 $\|A-UB\|_2$ 达到极小，如前，计算

$$\|A-UB\|_2^2=[A-UB,A-UB]=\|A\|_2^2-2\mathrm{Re}[A,UB]+\|B\|_2^2.$$

于是，必须求使 $\mathrm{Re}[A,UB]=\mathrm{Re}\operatorname{tr}AB^*U^*$ 为极大的酉矩阵 U. 如果 $AB^*=V\Sigma W^*$ 是 AB 的奇异值分解. 则

$$\operatorname{Re}\operatorname{tr} AB^*U^* = \operatorname{Re}\operatorname{tr} V\Sigma W^*U^* = \operatorname{Re}\operatorname{tr} \Sigma W^*U^*V$$
$$= \operatorname{Re}\sum_{i=1}^{m}\sigma_i(AB^*)t_{ii},$$

其中 $T=[t_{ij}]=W^*U^*V$ 是酉矩阵. 当所有 $t_{ii}=1$，即当 $U=VW^*$ 时，这个和取极大；VW^* 正好是 AB^* 的极分解的酉部分.

因此，用形如 UB 的矩阵对 $A\in M_{m,n}$的最佳最小二乘逼近由 $UB=(VW^*)B$ 给出，其中 $B\in M_{m,n}$，而 $U\in M_m$ 是酉矩阵，$AB^*=V\Sigma W^*$ 是 AB^* 的奇异值分解，或 $AB^*=P(VW^*)$是 AB^* 的极分解；我们并不需要分别知道 V 和 W. 这个逼近的误差由

$$\min\{\|A-UB\|_2 : U\in M_m \text{ 是酉矩阵}\} = \|A-(VW^*)B\|_2$$
$$= \left[\|A\|_2^2+\|B\|_2^2-2\sum_{i=1}^{m}\sigma_i(AB^*)\right]^{1/2}$$

给出，其中$\{\sigma_i(AB^*)\}$是 AB^* 的奇异值的集合.

如果想知道 A 是否恰好是 B 的转动，则一个明显的必要条件是 $\|A\|_2=\|B\|_2$，而其必要充分条件是

$$\|A\|_2^2=\|B\|_2^2=\sum_{i=1}^{m}\sigma_i(AB^*),$$

其中$\{\sigma_i(AB^*)\}$是 AB^* 的奇异值的集合.

最后，如果考虑 $m=n$ 及 $B=I$ 的特殊情形，便有以下事实：用酉矩阵 $U\in M_n$对给定矩阵 $A\in M_n$ 的最佳最小二乘逼近由 $U=VW^*$ 给出，其中 $A=V\Sigma W^*$ 是 A 的奇异值分解或 $A=PU=P(VW^*)$是 A 的极分解；逼近的误差是

431

$$\|A-VW^*\|_2^2=\|A\|_2^2+\|I\|_2^2-2\sum_{i=1}^{n}\sigma_i(A)$$
$$=\sum_{i=1}^{n}\sigma_i^2(A)+n-2\sum_{i=1}^{n}\sigma_i(A)=\sum_{i=1}^{n}(\sigma_i(A)-1)^2,$$

其中$\{\sigma_i(A)\}$是 A 的奇异值的集合.

如同上例中的讨论部分，在所有酉矩阵 U 上求使 $\operatorname{Re}\operatorname{tr}AU$ 为极大的问题的解. 为了以后参考方便，把这个结果概括成下述定理.

7.4.9 定理 设 $A\in M_n$ 是给定的矩阵，且 $A=V\Sigma W^*$ 是 A 的奇异值分解. 那么，(a)问题

$$\max\{\operatorname{Re}\operatorname{tr} AU : U\in M_n \text{ 是酉矩阵}\}$$

有解 $U=WV^*$，且极大值是 $\sigma_1(A)+\cdots+\sigma_n(A)$，其中$\{\sigma_i(A)\}$是 A 的奇异值集合. (b)存在酉矩阵 $U\in M_n$ 使得 $AU\in M_n$ 是半正定 Hermite 矩阵. 酉矩阵 U 是使(a)中问题达到极大的矩阵，当且仅当 AU 是半正定矩阵；如果 A 是非奇异矩阵，则 U 是唯一确定的. AU 的特征值是 A 的奇异值.

证明： 计算

$$\operatorname{Re}\operatorname{tr} AU=\operatorname{Re}\operatorname{tr} V\Sigma W^*U=\operatorname{Re}\operatorname{tr}\Sigma(W^*UV)=\sum_{i=1}^{n}\operatorname{Re}\sigma_i(W^*UV)_{ii},$$

它只有当所有$(W^*UV)_{ii}=1$ 时才取极大值. 因为 W^*UV 是酉矩阵，这又当且仅当 $W^*UV=I$

或 $U=WV^*$ 时才成立. 对于 U 的这个选择, $AU=V\Sigma W^*WV^*=V\Sigma V^*$, 又因为 $\Sigma=\text{diag}(\sigma_1, \cdots, \sigma_n)$ 且所有 $\sigma_i\geq 0$, 所以 AU 是半正定 Hermite 矩阵. 如果 $U_1\in M_n$ 是使 AU_1 为半正定矩阵的任一酉矩阵, 因为奇异值是酉不变的, 所以 AU_1 的特征值是 A 的奇异值. 对 A 是非奇异的情形, U 的唯一性可由(7.3.3)的唯一性部分推出. □

对于任意矩阵 $A\in M_{m,n}$, AA^* 以及 A^*A 都是半正定矩阵, 并且 $\text{tr}\, AA^*=\text{tr}\, A^*A=\sigma_1^2(A)+\cdots+\sigma_{\min(m,n)}^2(A)$, 因 A^* 与 A 有相同的奇异值, 所以 $\text{tr}\, AA^*$ 可看作 A 和 A^* 的相应奇异值的乘积之和. 这个简单的论断可以推广到任意一对矩阵 A 和 B, 只要乘积 AB 和 BA 有定义且为半正定矩阵. 这个结果对研究几种类型的矩阵最优化问题有用. 432

7.4.10 定理 设 $A\in M_{m,n}$, $B\in M_{n,m}$, 且 $q=\min\{m, n\}$. 设 $\sigma_1(A), \cdots, \sigma_q(A)$ 和 $\sigma_1(B), \cdots, \sigma_q(B)$ 分别表示 A 和 B 按递减顺序排列的奇异值. 如果 $AB\in M_m$ 和 $BA\in M_n$ 是半正定矩阵. 则存在整数 1, 2, …, q 的一个排列 τ, 使得

$$\text{tr}\, AB=\text{tr}\, BA=\sum_{i=1}^{q}\sigma_i(A)\sigma_{\tau(i)}(B). \tag{7.4.11}$$

证明: 如果 $m=n$, 且 A 和 B 都是半正定矩阵, 又如果 A 与 B 可交换, 则它们可以同时酉对角化为 $A=U\Lambda U^*$ 和 $B=UMU^*$, 其中, $U\in M_m$ 是酉矩阵, $\Lambda=\text{diag}(\lambda_1, \cdots, \lambda_m)$, $M=\text{diag}(\mu_1, \cdots, \mu_m)$, 且所有 λ_i, μ_i 都是非负的. 这时, 有

$$\text{tr}\, AB=\text{tr}(U\Lambda U^*)(UMU^*)=\text{tr}\, U\Lambda MU^*=\text{tr}\,\Lambda M=\sum_{i=1}^{m}\lambda_i\mu_i.$$

因为特征值 λ_i, μ_i 也是 A 和 B 的奇异值, 所以在 $m=n$ 的特殊情形, 定理得证.

不失一般性, 假定 $m\leq n$, 因为, 如果 $m>n$, 只要在定理的叙述中互换 A 和 B 就可以了.

为了证明定理的一般情形, 只须证明, 对任意一对使得 $m\leq n$ 以及 AB 和 BA 都是半正定矩阵的 $A\in M_{m,n}$ 和 $B\in M_{n,m}$, 存在酉矩阵 $V\in M_n$ 和有标准正交行的矩阵 $Y\in M_{m,n}$ 使得变换

$$\hat{A}=Y^*AV \quad 和 \quad \hat{B}=V^*BY \tag{7.4.12}$$

能得到一对可交换的 $n\times n$ 半正定矩阵 $\hat{A}$ 和 $\hat{B}$. 在这种情形, 由上述结论可知

$$\text{tr}\, AB=\text{tr}\, ABYY^*=\text{tr}\, Y^*ABY=\text{tr}(Y^*AV)(V^*BY)$$

$$=\sum_{i=1}^{m}\sigma_i(Y^*AV)\sigma_{\tau(i)}(V^*BY)=\sum_{i=1}^{m}\sigma_i(\hat{A})\sigma_{\tau(i)}(\hat{B}).$$

注意到 $\hat{A}^*\hat{A}=V^*A^*YY^*AV=V^*A^*AV=(AV)^*(AV)$, 于是, $\hat{A}$ 的奇异值与 AV 相同, 因为 $(AV)(AV)^*=AA^*$, 所以它又与 A 的奇异值相同. 同理可证 $\hat{B}$ 的奇异值与 B 相同, 由此可得

$$\text{tr}\, AB=\sum_{i=1}^{m}\sigma_i(A)\sigma_{\tau(i)}(B)$$

433

这正是所要求的. 现在分三步来证明, 存在形如(7.4.12)的变换, 且具有所要求的性质.

(1)设 A 和 B 适合定理的假设. 由(1.3.20)可知, BA 的特征值与 AB 的相同(重特征值按重数计算), 再加上 $n-m$ 个零特征值. 如果 $\lambda_1, \cdots, \lambda_m$ 是 AB 的特征值, 且 $\Lambda=\text{diag}(\lambda_1, \cdots, \lambda_m)$, 因为由假定可知, AB 和 BA 都是 Hermite 矩阵, 所以有酉矩阵 $U\in M_m$ 和 $V\in M_n$ 使得

$$AB = U\Lambda U^* \quad 和 \quad BA = V\begin{bmatrix}\Lambda & 0\\ 0 & 0\end{bmatrix}V^*.$$

若 V 写成分块形式 $V=[V_1 \ V_2]$，其中 $V_1 \in M_{n,m}$ 且 $V_2 \in M_{n,n-m}$，则 V_1 是具有标准正交列的矩阵，因而 $V_1^* V_1 = I \in M_m$. 于是 $\Lambda = U^* ABU$ 且 $BA = V_1 \Lambda V_1^*$，故 $BA = (V_1 U^*)AB(UV_1^*)$. 设 $Y = UV_1^* \in M_{m,n}$，并且注意到 $YY^* = UV_1^* V_1 U^* = UU^* = I$，则 Y 有标准正交行且 $BA = Y^* ABY$. 令 $\hat{A} \equiv Y^* A \in M_n$ 和 $\hat{B} \equiv BY \in M_n$，算出 $\hat{A}\hat{B} = Y^* ABY = BA$ 及 $\hat{B}\hat{A} = BYY^* A = BA$；根据假定，乘积 BA 是半正定矩阵. 因此有形如(7.4.12)的变换(其中 $V=I$)，它使我们得到一对可交换的 $n \times n$ 矩阵，且它们的乘积是半正定矩阵. 但是，单项 $\hat{A}$ 和 $\hat{B}$ 可能不是半正定矩阵；不过，我们可以进一步要求形如(7.4.12)的变换达到这个条件.

(2) 不失一般性，现在可以假定，$m=n$，A，$B \in M_n$ 可交换，且乘积 AB 是半正定矩阵. 如果 $(AB)x = \lambda x$，且 $x \neq 0$，则 $(AB)(Ax) = ABAx = AABx = A(ABx) = A\lambda x = \lambda(Ax)$，所以 Hermite 矩阵 AB 的每个特征空间在 A 作用下不变. 同理可证，这每一个特征空间也在 B 的作用下不变. 因此，如果 $U = [u_1 \cdots u_n]$ 是以 AB 的特征向量为列构成的酉矩阵，且相应于 AB 的同一特征值的所有特征向量相邻地排放在一起，则 $U^* AU$ 和 $U^* BU$ 一定都是分块对角矩阵，且

$$\hat{A} = U^* AU = \mathrm{diag}(A_1, A_2, \cdots, A_r), \quad \hat{B} = U^* BU = \mathrm{diag}(B_1, B_2, \cdots, B_r),$$

其中，A_i，$B_i \in M_{k_i}$，$1 \leqslant k_i \leqslant n$，$k_1 + \cdots + k_r = n$，且每个 $A_i B_i = B_i A_i = \lambda_i I \in M_{k_i}$，另外 λ_1，λ_2，$\cdots$，λ_r 是半正定矩阵 AB 的不同(非负)特征值.

(3)不失一般性，现在可以假定，$m=n$，A，$B \in M_n$ 可交换，$AB = \lambda I$，且 $\lambda \geqslant 0$. 如果 $\lambda > 0$，则 A 和 B 都是非奇异矩阵，且 $B = \lambda A^{-1}$. 利用(7. 4. 9)求酉矩阵 $U \in M_n$ 使得 $\hat{A} \equiv AU$ 是半正定矩阵. 另一方面，因为 $\lambda > 0$，所以 $\hat{B} \equiv U^* B = \lambda U^* A^{-1} = \lambda (AU)^{-1}$ 也是半正定矩阵，因而 $(AU)^{-1}$ 是半正定矩阵. 此外，$(U^* B)(AU) = U^* \lambda IU = \lambda I = AB = (AU)(U^* B)$，所以 $\hat{A}$ 与 $\hat{B}$ 可交换. 这是形如(7. 4. 12)的变换，因此，如果 $\lambda > 0$，就完成了证明.

[434]

如果 $\lambda = 0$，则 $AB = BA = 0$. 再选择酉矩阵 U 使得 AU 是半正定矩阵. 于是 $0 = AB = (AU)(U^* B) = (U^* B)(AU) = U^* 0U = 0$，因而 AU 和 $U^* B$ 可交换，且 Hermite 矩阵 AU 的每个特征空间在 $U^* B$ 作用下不变，若 $W = [w_1 \cdots w_n]$ 是以 AU 的特征向量为列构成的酉矩阵，且相应于 AU 的同一特征值的所有特征向量相邻地排放在一起，则 $W^*(AU)W$ 和 $W^*(U^* B)W$ 都是分块对角矩阵，且

$$W^*(AU)W = \mathrm{diag}(A_1, \cdots, A_r), W^*(U^* B)W = \mathrm{diag}(B_1, \cdots, B_r),$$

A_i 和 B_i 是同阶矩阵，而 $A_i = \lambda_i I$，$i=1$，2，$\cdots$，r，其是 λ_1，λ_2，$\cdots$，λ_r 是半正定矩阵 AU 的不同(非负)特征值. 对所有 $i=1$，$\cdots$，r，我们有 $A_i B_i = B_i A_i = 0$. 如果 $\lambda_i \neq 0$，则 $A_i = \lambda_i I$ 和 $B_i = 0$ 是一对可交换的半正定矩阵，这正是所要求的. 如果 $\lambda_i = 0$，则 B_i 不一定是零矩阵，但是，存在酉矩阵 U_i 使得 $U_i^* B_i$ 是半正定矩阵[应用(7.4.9)于 B^*]，因而，在这种情形，$A_i U_i = 0$ 和 $U_i^* B_i$ 构成一对由形如(7.4.12)的变换得到的可交换半正定矩阵. 这样，对所有可能情形都进行了验证.

□

7.4.13 例 作为(7.4.8)中旋转问题的变形，设 A，$B \in M_{m,n}$ 是给定的矩阵，并且我们希望确

定，是否可以通过 B 的两侧"旋转"来得到 A；也就是 $A=UBV$ 对某两个酉矩阵 $U\in M_m$，$V\in M_n$ 成立吗？更一般地，如果考虑已知矩阵 B 的所有可能的两侧"旋转"UBV，在最小二乘意义下，可以怎样充分地逼近 A？

如前，我们试图选择使 $\|A-UBV\|_2$ 为极小的酉矩阵 $U\in M_m$ 和 $V\in M_n$，还和前面一样，算出

$$\|A-UBV\|_2^2=[A-UBV,A-UBV]=\|A\|_2^2-2\mathrm{Re}[A,UBV]+\|B\|_2^2.$$

因此，必须求使 $\mathrm{Re}[A,UBV]=\mathrm{Re}\,\mathrm{tr}AV^*B^*U^*$ 为极大的酉矩阵 $U\in M_m$ 和 $V\in M_n$。使这个问题有极大值的酉矩阵 U_0，V_0 一定存在(但不一定唯一)，这是因为 M_n 中和 M_m 中酉矩阵的集合是紧集，又紧集的笛卡儿乘积也是紧集。使上述问题为极大的矩阵 U_0，V_0 对任意酉矩阵 $U\in M_m$ 有性质

$$\mathrm{Re}\,\mathrm{tr}(AV_0^*B^*)U_0^*\geqslant\mathrm{Re}\,\mathrm{tr}(AV_0^*B^*)U,$$

于是，由(7.4.9)可知，$AV_0^*B^*U_0^*$ 是半正定矩阵，同理可证，对任意酉矩阵 $V\in M_n$，

$$\mathrm{Re}\,\mathrm{tr}\,AV_0^*B^*U_0^*=\mathrm{Re}\,\mathrm{tr}(B^*U_0^*A)V_0^*\geqslant\mathrm{Re}\,\mathrm{tr}\,B^*U_0^*AV$$

于是，仍由(7.4.9)可知，$B^*U_0^*AV_0^*$ 是半正定矩阵，这样，两个矩阵 $AV_0^*\in M_{m,n}$和 $B^*U_0^*\in M_{n,m}$都适合定理(7.4.10)的假设，因为奇异值是酉不变的，因此，如果 $q=\min\{m,n\}$，则对整数 1，…，q 的某个置换 τ，有

$$\max\{\mathrm{Re}\,\mathrm{tr}AV^*B^*U^*:U\in M_m\text{ 和 }V\in M_n\text{ 是酉矩阵}\}$$

$$=\mathrm{Re}\,\mathrm{tr}AV_0^*B^*U_0^*=\sum_{i=1}^{q}\sigma_i(AV_0^*)\sigma_{\tau(i)}(B^*U_0^*)=\sum_{i=1}^{q}\sigma_i(A)\sigma_{\tau(i)}(B).$$

不失一般性，我们把奇异值 $\sigma_1(A)$，…，$\sigma_q(A)$和 $\sigma_1(B)$，…，$\sigma_q(B)$按递减顺序排列．如果置换 τ 不是恒等置换，则有适合 $1\leqslant i_1<i_2\leqslant q$ 的指标使$\sigma_{\tau(i_1)}(B)\leqslant\sigma_{\tau(i_2)}(B)$，并且容易验证，如果所作置换互换这两个奇异值的位置，则和

$$\sum_{i=1}^{q}\sigma_i(A)\sigma_{\tau(i)}(B)$$

不会减小，事实上，新和值与旧和值之差是

$$[\sigma_{i_1}(A)-\sigma_{i_2}(A)][\sigma_{\tau(i_2)}(B)-\sigma_{\tau(i_1)}(B)]\geqslant 0.$$

因此，对恒等置换 τ，上述和达到极大值，并且可以得出

$$\max\{\mathrm{Re}\,\mathrm{tr}AV^*B^*U^*:U\in M_m,V\in M_n\text{ 是酉矩阵}\}=\sum_{i=1}^{q}\sigma_i(A)\sigma_i(B),\quad(7.4.14)$$

其中 A 和 B 的奇异值都按递减顺序排列．

把这个结果用到当初求极小值的问题中，对 A，$B\in M_{m,n}$，$q=\min\{m,n\}$，求得

$$\min\{\|A-UBV\|_2:U\in M_m\text{ 和 }V\in M_n\text{ 是酉矩阵}\}$$

$$=\Big[\|A\|_2^2-2\sum_{i=1}^{q}\sigma_i(A)\sigma_i(B)+\|B\|_2^2\Big]^{1/2}$$

$$=\Big[\sum_{i=1}^{q}\sigma_i^2(A)-2\sum_{i=1}^{q}\sigma_i(A)\sigma_i(B)+\sum_{i=1}^{q}\sigma_i^2(B)\Big]^{1/2}$$

436

$$= \left[\sum_{i=1}^{q} [\sigma_i(A) - \sigma_i(B)]^2 \right]^{1/2}. \quad (7.4.15)$$

特别是，A 是 B 的两侧“旋转”，当且仅当 A 和 B 有相同的奇异值集合. □

练习 如果 $B=I$，(7.4.15)说明什么？试与例(7.4.8)末尾的结果相比较. 如果 B 是秩 k 的对角矩阵，(7.4.15)说明什么？试与例(7.4.1)中的有关说明作一比较.

7.4.16 例 作为利用奇异值的另一个例子，我们考虑刻划矩阵的酉不变范数的问题，它是在(5.6)节中提出来的.

定义 $M_{m,n}$上的向量范数 $\|\cdot\|$ 称为酉不变的，是指

$$\|UAV\| = \|A\|$$

对所有 $A\in M_{m,n}$ 和所有酉矩阵 $U\in M_m$，$V\in M_n$ 成立.

如果 $A\in M_{m,n}$ 是给定的矩阵，且 $A=V\Sigma W^*$ 是 A 的奇异值分解，则 $\|A\| = \|V\Sigma W^*\| = \|\Sigma\|$ 对任意酉不变范数 $\|\cdot\|$ 成立. 因此，某个阶数的矩阵的酉不变范数只与该矩阵的奇异值集合有关.

两个熟知的酉不变范数的例子是 Frobenius(Euclid)范数和谱范数. 如果 $X=[x_{ij}]\in M_{m,n}$ 的奇异值是 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_q\geqslant 0$ $(q=\min\{m, n\})$，则

$$\|X\|_2 = \left(\sum_{j=1}^{n}\sum_{i=1}^{m} |x_{ij}|^2\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{q}\sigma_i^2\right)^{1/2},$$

且

$$\|X\|_2 = \max_{y\neq 0} \frac{\|Xy\|_2}{\|y\|_2} = [\rho(X^*X)]^{1/2} = \sigma_1 = \max\{\sigma_1, \cdots, \sigma_q\}.$$

对于 $M_{m,n}$ 上的一般酉不变范数 $\|\cdot\|$，它对其自变量的奇异值的依赖关系是容易确定的. 为方便起见，假定 $m\leqslant n$，设 $A=\operatorname{diag}(x_1, x_2, \cdots, x_m)\in M_m$，然后定义分块矩阵

$$X = [A|0], \quad A\in M_m, \quad 0\in M_{m,n-m}.$$

因为 $XX^*=\operatorname{diag}(|x_1|^2, |x_2|^2, \cdots, |x_m|^2)$，所以 X 的奇异值集合是 $\{\sigma_i\}=\{|x_i|\}$. 如果定义函数 g：$\mathbf{C}^m\to\mathbf{R}^1$ 为

$$g(x) = g([x_1, \cdots, x_m]^T) = \|X\|,$$

则函数 $g(\cdot)$ 从范数 $\|\cdot\|$ 继承了某些性质：

(7.4.17) $g(x)\geqslant 0$ 对所有 $x\in\mathbf{C}^m$ 成立，因为 $\|X\|\geqslant 0$ 对所有 $X\in M_{m,n}$ 成立.

437

(7.4.18) $g(x)=0$ 当且仅当 $x=0$，因为 $\|X\|=0$ 当且仅当 $X=0$.

(7.4.19) $g(\alpha x)=|\alpha|g(x)$ 对所有 $x\in\mathbf{C}^m$ 和所有 $\alpha\in\mathbf{C}$ 成立，因为 $\|\alpha X\| = |\alpha|\|X\|$ 对所有 $\alpha\in\mathbf{C}$ 和所有 $X\in M_{m,n}$ 成立.

(7.4.20) $g(x+y)\leqslant g(x)+g(y)$ 对所有 x，$y\in\mathbf{C}^m$ 成立，因为 $\|X+Y\|\leqslant\|X\|+\|Y\|$ 对所有 X，$Y\in M_{m,n}$ 成立.

这四个性质说明 $g(\cdot)$ 一定是 $\mathbf{C}^m$ 上的向量范数，但是 $g(\cdot)$ 另有两个附加性质：

(7.4.21) 正如(5.5.9)中所定义的那样，$g(\cdot)$ 是 $\mathbf{C}^m$ 上的绝对范数；也就是说，如果 $x=[x_i]\in\mathbf{C}^m$，且 $y=[y_i]\equiv[|x_i|]\in\mathbf{C}^m$，则 $g(x)=g(y)$. 这是因为 $g(x)=\|X\|$ 只与 X

上的奇异值$\sigma_i = |x_i|$有关.

(7.4.22) 如果$P\in M_m$是置换矩阵，则$g(Px)=g(x)$对所有$x\in \mathbf{C}^m$成立，这是因为$X=[A\,|\,0]$的奇异值集合与$[PA\ \ 0]$相同，这又因为$(PA)^*(PA)=A^*P^TPA=A^*A$. 函数$g(x)$只是x的各分量绝对值的集合的函数，无需考虑它们的顺序.

练习 试计算矩阵$X=[A\,|\,0]\in M_{m,n}$的一个明显的奇异值分解，其中$A=\operatorname{diag}(x_1,\cdots,x_m)$，刚才讨论了这样的奇异值分解.

练习 如果$m\geqslant n$. 设$X=[A\,|\,0]^T$，其中$A=\operatorname{diag}(x_1,\cdots,x_n)\in M_n$，且定义$g(x)=\|X\|$，$x\in\mathbf{C}^n$. 如果$\|\cdot\|$是$M_{m,n}$上的酉不变范数，证明，$g(\cdot)$是$\mathbf{C}^n$上的绝对向量范数，且$g(Px)=g(x)$对所有$x\in\mathbf{C}^n$和每个置换矩阵$P\in M_n$成立.

练习 直接证明，由 Frobenius 范数和谱范数诱导的向量范数适合上述(7.4.17)～(7.4.22)六个性质.

7.4.23 定义 函数$g(\cdot)$: $\mathbf{C}^q\to\mathbf{R}^+$称为对称度规函数，当且仅当它适合上述(7.4.17)～(7.4.22)六个性质，即当且仅当$g(\cdot)$是绝对向量范数且$g(\cdot)$是其自变量的诸坐标的置换不变函数.

我们已经看到，$M_{m,n}$上的每个酉不变范数诱导一个对称度规函数；更有趣的是其逆命题也成立. 下面的定理说明，$M_{m,n}$上的函数$N(\cdot)$是酉不变范数，当且仅当$N(A)$是A的奇异值的对称度规函数.

7.4.24 定理 设$\|\cdot\|$是$M_{m,n}$上的酉不变范数，并且$q=\min\{m,n\}$，设$x=[x_i]\in\mathbf{C}^q$和$X_1=\operatorname{diag}(x_1,\cdots,x_q)$，当$m\leqslant n$时，设$X\equiv[X_1\ \ 0]\in M_{m,n}$，或当$m\geqslant n$时设$X\equiv[X_1\,|\,0]^T\in M_{m,n}$. 设$g$: $\mathbf{C}^q\to\mathbf{R}^+$定义为$g(x)\equiv\|X\|$. 则$g(\cdot)$是对称度规函数. 反之，如果$g$: $\mathbf{C}^q\to\mathbf{R}^+$是给定的对称度规函数，又如果$\|\cdot\|$: $M_{m,n}\to\mathbf{R}^+$用$\|A\|=g([\sigma_1,\cdots,\sigma_q]^T)$来定义，其中$\sigma_1,\cdots,\sigma_q$是$A$的奇异值，则$\|\cdot\|$是$M_{m,n}$上的酉不变范数. [438]

证明：前一个论断已经证明. 关于逆命题，我们注意到，因为$g(\cdot)$是其变量的诸分量的置换不变函数，所以$\|\cdot\|$是$M_{m,n}$上有明确定义的函数. 由于矩阵的奇异值集合是酉不变的，所以，对于所有酉矩阵$U\in M_m$和$V\in M_n$，还有$\|UAV\|=\|A\|$. 因为$g(\cdot)$是向量范数，所以，对所有$A\in M_{m,n}$都有$\|A\|\geqslant 0$，并且$\|A\|=0$当且仅当$g([\sigma_1,\cdots,\sigma_q]^T)=0$，而这又当且仅当所有$\sigma_i=0$才能成立. 这是因为$g(\cdot)$是正定的(7.4.18). 但是零矩阵是所有奇异值都是零的仅有矩阵，因而函数$\|\cdot\|$是正定的(见 5.1.1(1a)). 因为$\sigma_i(cA)=|c|\sigma_i(A)$，所以$\|cA\|=g([|c|\sigma_1,\cdots,|c|\sigma_q]^T)=|c|g([\sigma_1,\cdots,\sigma_q]^T)=|c|\|A\|$，因而$\|\cdot\|$也是齐次的. 至此，已经证明了，用对称度规函数这种方式诱导的任意函数$\|\cdot\|$是$M_{m,n}$上的准范数[见(5.4)]. 剩下要证明的是$\|\cdot\|$适合三角不等式，为此，只需证明$\|\cdot\|$是准范数的对偶范数，[参看(5.4.12)下面的讨论]因而它实际上也是范数.

考虑$\mathbf{C}^q$上范数$g(\cdot)$的对偶范数$g^D(\cdot)$：

$$g^D(y)\equiv\max_{g(x)=1}\operatorname{Re} y^*x \tag{7.4.25}$$

因为$g(\cdot)$是(准)范数，所以函数$g^D(\cdot)$一定是范数；因为$g(\cdot)$适合(7.4.21)，所以，

(7.4.21′) 如果 $E=\mathrm{diag}(e^{i\theta_1},\cdots,e^{i\theta_q})$，且所有 $\theta_i\in\mathbf{R}$，则

$$g^D(Ey)=\max_{g(x)=1}\mathrm{Re}\,(Ey)^*x=\max_{g(x)=1}\mathrm{Re}\,y^*(E^*x)=\max_{g(E^*x)=1}\mathrm{Re}\,y^*x$$

$$=\max_{g(x)=1}\mathrm{Re}\,y^*x=g^D(y).$$

因此 $g^D(\cdot)$也是对称度规函数．从而 $g^D(\cdot)$也适合(7.4.21).

(7.4.22′) 同理可证，如果 $P\in M_q$ 是置换矩阵．因为 $g(\cdot)$适合(7.4.22)，所以

$$g^D(Py)=\max_{g(x)=1}\mathrm{Re}\,(Py)^*x=\max_{g(x)=1}\mathrm{Re}\,y^*P^Tx=\max_{g(Px)=1}\mathrm{Re}\,y^*x$$

439

$$=\max_{g(x)=1}\mathrm{Re}\,y^*x=g^D(y).$$

这样，我们可以在 M_n 上定义与对称度规函数 $g^D(\cdot)$相关联的函数 $\|\cdot\|^D$：

$$\|A\|^D\equiv g^D([\sigma_1,\cdots,\sigma_q]^T),$$

其中 $\sigma_1,\cdots,\sigma_q$ 是 A 的奇异值．[这里，有意泛用了一个记号：$\|\cdot\|^D$ 通常表示范数 $\|\cdot\|$ 的对偶范数；尽管还不知道 $\|\cdot\|$ 是范数，不过将证明 $\|\cdot\|$ 是范数，并且证明，当 $\|\cdot\|^D$ 用对称度规范数定义时，它就是对偶范数.]我们已经证明这个函数 $\|\cdot\|^D$ 是 M_q 上的准范数，因为它是用对称度规函数 $g^D(\cdot)$定义的.

现在计算 $\|\cdot\|^D$ 的对偶，根据(5.4.12)，$\|\cdot\|^D$ 肯定是 $M_{m,n}$上的范数．我们知道，矩阵 $B\in M_{m,n}$适合 $\|B\|^D=1$，当且仅当 B 的奇异值分解是 $B=V\Sigma W^*$[其中，酉矩阵 $V\in M_m$ 和 $W\in M_n$，$\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_q)$]且 $g^D([\sigma_1,\cdots,\sigma_q]^T)=1$．对于每个给定的矩阵 $A\in M_{m,n}$，有

$$(\|A\|^D)^D=\max_{\|B\|^D=1}\mathrm{Re}[A,B]=\max_{\|B\|^D=1}\mathrm{Re}\,\mathrm{tr}AB^*$$

$$=\max\{\mathrm{Re}\ \mathrm{tr}\ A(V\Sigma W^*)^*:V\in M_m \text{ 和 } W\in M_n \text{ 是酉矩阵},$$

$$\Sigma=\mathrm{diag}(s_1,\cdots,s_q),\text{且}$$

$$g^D([s_1,\cdots,s_q]^T)=1\}.$$

对于适合上述约束条件的每个对角矩阵 Σ，我们可以利用(7.4.14)计算这个极大值，并且能在酉矩阵 V，W 的所有选择上达到极大值：

$$(\|A\|^D)^D=\max\Big\{\sum_{i=1}^{q}\sigma_i(A)\,|s_i|:g^D([s_1,\cdots,s_q]^T)=1\Big\}.$$

但是，因为所有 $\sigma_i(A)\geqslant 0$，由定义(5.4.12)显然可知，这个最大值正好就是在点$[\sigma_1(A),\cdots,\sigma_q(A)]^T$ 取值的 $g^D(\cdot)$的对偶范数．然而，对偶定理(5.5.14)保证，范数的对偶的对偶是原范数，因而

$$(\|A\|^D)^D=(g^D)^D([\sigma_1(A),\cdots,\sigma_q(A)]^T)=g([\sigma_1(A),\cdots,\sigma_q(A)]^T)\equiv\|A\|.$$

于是，对所有 $A\in M_{m,n}$，$\|A\|=(\|A\|^D)^D$．它保证 $\|\cdot\|$ 实际上是范数，因而它适合三角不等式．这个结论也说明泛用这个记号是合理的，这是因为，由对偶定理可知，$(\|A\|)^D=((\|A\|^D)^D)^D=\|A\|^D$．因此，用对称度规函数 $g^D(\cdot)$定义的 $\|\cdot\|^D$ 确实与范数 $\|\cdot\|$ 的对偶范数相同.

□

440

$\mathbf{C}^n$ 上的对称度规函数的一个熟知的重要例子是 l_p 范数族(5.2.4)

$$g([x_1,\cdots,x_n]^T)=\|x\|_p=\Big(\sum_{i=1}^{n}|x_i|^p\Big)^{1/p},\quad 1\leqslant p<\infty.$$

当它应用于矩阵的奇异值时，正如定理(7.4.24)中所描述的，各种 l_p 范数诱导了 $M_{m,n}$ 上的各种酉不变范数，称为 Schatten p 范数，$p=2$ 的情形是 Frobenius(Euclidean)范数

$$\|A\|_2 = \left[\sum_i \sigma_i(A)^2\right]^{1/2}.$$

$p\to\infty$ 的极限情形是谱范数

$$\|A\|_2 = \max\{\sigma_i(A)\},$$

而 $p=1$ 的情形是迹范数

$$\|A\|_{tr} = \sum_i \sigma_i(A).$$

在例(7.4.6)中，当考虑用一个酉矩阵的纯量倍来逼近方阵的问题时自然要出现迹范数.

$\mathbf{C}^n$ 上的另一类对称度规函数族在(7.4.44)中给出，它们也诱导出迹范数和谱范数.

7.4.26 例 奇异值在推导 Wielandt 不等式中起着重要作用，这个不等式给出了非奇异方阵关于谱范数的条件数的几何意义.

设 $A\in M_n$ 是非奇异矩阵，设 $B=A^*A\in M_n$，且用 $\sigma_1\geqslant\cdots\geqslant\sigma_n>0$ 表示 A 的奇异值. 正定矩阵 B 的奇异值(按约定的递增顺序排列)是 $0<\sigma_n^2\leqslant\sigma_{n-1}^2\leqslant\cdots\leqslant\sigma_1^2$. 设 x，$y\in\mathbf{C}^n$ 是任意一对标准正交向量，定义 $C=[x\,y]^*B[x\,y]\in M_2$，且用 $0<\gamma_1\leqslant\gamma_2$ 表示 C 的特征值. $r=2$ 的 Poincaré 分离定理(4.3.16)说明

$$\lambda_k(B)=\sigma_{n-k+1}^2\leqslant\lambda_k(C)=\gamma_k\leqslant\lambda_{n+k-2}(B)=\sigma_{3-k}^2,\quad k=1,2,$$

或

$$\sigma_n^2\leqslant\gamma_1\leqslant\sigma_2^2 \quad 和 \quad \sigma_{n-1}^2\leqslant\gamma_2\leqslant\sigma_1^2.$$

对实际应用来说，在这些不等式中，值得注意的关系是

$$\sigma_n^2\leqslant\gamma_1\leqslant\gamma_2\leqslant\sigma_1^2, \tag{7.4.27}$$

其中，如果 x 和 y 是 B 的相应于其特征值分别是 A 的最大和最小奇异值的平方的标准正交特征向量，则第一个和最后一个不等式是等式.

经计算，

$$\begin{aligned}1-\frac{|x^*By|^2}{(x^*Bx)(y^*By)} &= 4\frac{(x^*Bx)(y^*By)-|x^*By|^2}{(x^*Bx+y^*By)^2-(x^*Bx-y^*By)^2}\\ &=\frac{4\det C}{(\operatorname{tr}C)^2-(x^*Bx-y^*By)^2}\\ &=\frac{4\gamma_1\gamma_2}{(\gamma_1+\gamma_2)^2-(x^*Bx-y^*By)^2}\geqslant\frac{4\gamma_1\gamma_2}{(\gamma_1+\gamma_2)^2},\end{aligned} \tag{7.4.28}$$

其中等式成立当且仅当 x，$y\in\mathbf{C}^n$ 是标准正交向量且 $x^*Bx=y^*By$. 我们把这个不等式变成等价的不等式

$$\frac{|x^*By|^2}{(x^*Bx)(y^*By)}\leqslant 1-\frac{4\gamma_1\gamma_2}{(\gamma_1+\gamma_2)^2}=\left(\frac{\gamma_1-\gamma_2}{\gamma_1+\gamma_2}\right)^2=\left(\frac{\gamma_2/\gamma_1-1}{\gamma_2/\gamma_1+1}\right)^2. \tag{7.4.29}$$

(7.4.29)中的上界是比值 γ_2/γ_1 的单增函数[这可用当 $t>0$ 时函数 $f(t)=(t-1)/(t+1)$ 的导数为正的事实来证明]. 根据(7.4.27)，这个比值有上界 σ_1^2/σ_n^2. 因而

$$\frac{|x^*By|^2}{(x^*Bx)(y^*By)}\leqslant\left(\frac{\sigma_1^2/\sigma_n^2-1}{\sigma_1^2/\sigma_n^2+1}\right)^2=\left(\frac{\kappa^2-1}{\kappa^2+1}\right)^2, \tag{7.4.30}$$

441

其中，引进了正参数 $\kappa=\kappa(A)=\sigma_1/\sigma_n=\|A\|_2\|A^{-1}\|_2\geqslant 1$，它是 A 关于谱范数的条件数，如果 u_1、$u_n\in\mathbf{C}^n$ 分别是相应于特征值 σ_1^2 和 σ_n^2 的标准正交特征向量，又如果 $x=(u_1+u_n)/\sqrt{2}$，$y=(u_1-u_n)/\sqrt{2}$，则 $\{x, y\}$ 是标准正交组，$x^*Bx=y^*By=(\sigma_1^2+\sigma_n^2)/2$，且 $x^*By=(\sigma_1^2-\sigma_n^2)/2$，这时(7.4.30)中取等号.

用 $\cot(\theta/2)=\kappa$ 定义第一象限中的角 θ，于是

$$\frac{\kappa^2-1}{\kappa^2+1}=\frac{\cot^2(\theta/2)-1}{\cot^2(\theta/2)+1}=\frac{\cos^2(\theta/2)-\sin^2(\theta/2)}{\cos^2(\theta/2)+\sin^2(\theta/2)}=\cos\theta,$$

并且(7.4.30)可写成形式

$$\frac{|x^*By|^2}{(x^*Bx)(y^*By)}\leqslant\cos^2\theta. \tag{7.4.31}$$

如果我们注意到这个不等式的左边关于 x 和 y 是零次齐次的，那么最后可以用两种等价的形式叙述 Wielandt 不等式：

7.4.32 定理 设 $A\in M_n$ 是具有谱条件数 κ 的某个非奇异矩阵，且用 $\cot(\theta/2)=\kappa$ 定义第一象限中的角 θ，则

[442]

$$|\langle Ax, Ay\rangle|\leqslant\cos\theta\|Ax\|_2\|Ay\|_2 \tag{7.4.33}$$

对每对正交向量 x，$y\in\mathbf{C}^n$ 成立，其中 $\langle u, v\rangle\equiv v^*u$ 表示 Euclid 内积，而 $\|u\|_2=(u^*u)^{1/2}$ 表示 Euclid 范数. 另外，存在一对标准正交向量 x，$y\in\mathbf{C}^n$ 使(7.4.33)中的等式成立.

7.4.34 定理 设 $B\in M_n$ 是具有特征值 $0<\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$ 的某个正定矩阵. 则

$$|x^*By|^2\leqslant\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^2(x^*Bx)(y^*By) \tag{7.4.35}$$

对每对正交向量 x，$y\in\mathbf{C}^n$ 成立. 此外，存在一对标准正交向量 x，$y\in\mathbf{C}^n$ 使得(7.4.35)中等式成立.

证明：在(7.4.31)中作代换 $B=A^*A$ 便得到不等式(7.4.33)，在(7.4.30)中作代换 $\sigma_i^2=\lambda_{n-i+1}$，并且考虑到每个正定矩阵 B 具有形式 $B=A^*A$，其中 $A\in M_n$ 为某个非奇异矩阵，由此得到不等式(7.4.35)；可以取 $A=B^{1/2}$. 我们已经知道(7.4.30)中的等式对一对标准正交向量成立. □

练习 证明，(7.4.35)是一般的 Cauchy-Schwarz 不等式的改进，它是 $|x^*By|=|\langle Cy, Cx\rangle|\leqslant\|Cx\|_2\|Cy\|_2$，其中 $C=B^{1/2}$. 不过 Cauchy-Schwarz 不等式适应于所有向量偶 x，y，而(7.4.35)只适用于正交向量偶. 如果 $\lambda_1=\lambda_n$，会出现什么情况？

Wielandt 不等式的形式(7.4.33)直接导出谱条件数的一个有用的几何解释. 如果 x，$y\in\mathbf{C}^n$ 是任意一对标准正交向量，则不等式

$$\frac{|\langle Ax, Ay\rangle|}{\|Ax\|_2\|Ay\|_2}\leqslant\cos\theta \tag{7.4.36}$$

左边是非零向量 Ax 与 Ay 间的较小的 Euclid 角的通常余弦. 这个界说明，Ax 与 Ay 间的较小角至多是 $\theta=\theta(A)$，其中 $\theta(A)$ 用 $\cot[\theta(A)/2]=\kappa(A)$ 来定义. 因为在这个界中可以取等式，所以已经证实，当 x 和 y 取遍所有可能的标准正交向量时，$\theta(A)$ 可几何地解释为 Ax 与 Ay 间的

最小角．这一论点已在(5.8)节和(6.3)节讨论过． 443

众所周知的 Kantorovich 不等式容易从 Wielandt 不等式推出．对任意 $x\in\mathbf{C}^n$，定义

$$y = \|x\|_2^2(B^{-1}x) - (x^*B^{-1}x)x, \tag{7.4.37}$$

并且注意到 $x^*y=0$．经计算，

$$By = \|x\|_2^2 x - (x^*B^{-1}x)Bx,$$

$$x^*By = \|x\|_2^4 - (x^*B^{-1}x)(x^*Bx),$$

$$y^*By = -(x^*B^{-1}x)(y^*Bx).$$

因为 B，因而 B^{-1} 都是正定矩阵，我们一定有 $y^*By\geqslant 0$，因而 $y^*Bx=x^*By\leqslant 0$．把不等式(7.4.31)写成形式

$$|x^*By|^2 \leqslant \cos^2\theta(x^*Bx)(y^*By).$$

然后特意选择适合(7.4.37)的一对 x，y 代入上式便得

$$|x^*By|^2 \leqslant (\cos^2\theta)(x^*Bx)(x^*B^{-1}x)(-x^*By).$$

在 $x^*By<0$ 或 $x^*By=0$ 两种可能情形，这蕴涵，对任意 $x\in\mathbf{C}^n$，有

$$-x^*By = -[\|x\|_2^4 - (x^*B^{-1}x)(x^*Bx)] \leqslant (\cos^2\theta)(x^*Bx)(x^*B^{-1}x),$$

或

$$(\sin^2\theta)(x^*Bx)(x^*B^{-1}x) \leqslant \|x\|_2^4. \tag{7.4.38}$$

应指出的是，如果 $x=u_1+u_n$ 是 B 的相应于它的最小特征值和最大特征值的单位正交特征向量的和，则(7.4.38)是等式．这就诱导出 Kantorovich 不等式的两种形式，它们与 Wielandt 不等式的两种形式相对应．

7.4.39 定理 设 $A\in M_n$ 是具有谱条件数 κ 的某个非奇异矩阵，且用 $\cot(\theta/2)=\kappa$ 定义第一象限的角 θ．则

$$\|x\|_2^2 \geqslant \sin\theta\|Ax\|_2\|(A^*)^{-1}x\|_2 \tag{7.4.40}$$

对所有 $x\in\mathbf{C}^n$ 成立．此外，存在单位向量 x 使(7.4.40)为等式．

7.4.41 定理 设 $B\in M_n$ 是具有特征值 $0<\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$ 的某个正定矩阵．则

$$\|x\|_2^4 \geqslant \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}(x^*Bx)(x^*B^{-1}x) \tag{7.4.42}$$

444

对所有 $x\in\mathbf{C}^n$ 成立．此外，存在单位向量 x 使(7.4.42)为等式．

证明： 将 $B=A^*A$ 代入(7.4.38)，并且能想到

$$\sin^2\theta = 1-\cos^2\theta = 1-\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^2 = \frac{4\lambda\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}$$

便可从(7.4.38)推出这两个结论．(7.4.40)和(7.4.42)中可能取等式的事实可以从(4.4.38)中取等式的情形推出． □

7.4.43 例 有时可能要证明某些对所有酉不变范数都成立的关于矩阵的范数不等式，证明的关键在于认识到用

$$g_k(x) = \max\{|x_{i_1}|+\cdots+|x_{i_k}|: 1\leqslant i_1<i_2<\cdots<i_k\leqslant n\},\quad k=1,\cdots,n. \tag{7.4.44}$$

定义的 $\mathbf{C}^n$ 上的特殊对称度规函数 $g_k([x_1, \cdots, x_n]^T)$ 的基本作用. 当把它应用于矩阵的奇异值时，正如定理(7.4.24)中所描述的，这个特殊的对称度规函数族诱导出 $M_{m,n}$ 上的酉不变范数族，称之为樊畿 k 范数. $k=1$ 的情形是谱范数，而 $k=\min\{m, n\}$ 的情形是迹范数.

7.4.45 定理 设 $x=[x_i]$，$y=[y_i]\in\mathbf{C}^n$ 是给定的向量，则 $g(x)\leqslant g(y)$ 对 $\mathbf{C}^n$ 上的所有对称度规函数 $g(\cdot)$ 成立，当且仅当 $g_k(x)\leqslant g_k(y)$ 对 $k=1, 2, \cdots, n$ 成立；其中 $g_k(\cdot)$ 是(7.4.44)中所定义的特殊对称度规函数.

证明： 因为每个 $g_k(\cdot)$ 是对称度规函数，条件的必要性是显然的. 为了证明充分性，假定 $g_k(x)\leqslant g_k(y)$ 对 $k=1, 2, \cdots, n$ 成立，且设 $g(\cdot)$ 是给定的对称度规函数. 由于对称度规函数是其自变量的分量的置换不变函数(7.4.22)，为方便起见，不失一般性，我们可以假定 x 和 y 的分量的绝对值都排成递增顺序：

$$|x_1|\leqslant|x_2|\leqslant\cdots\leqslant|x_n|,\quad |y_1|\leqslant|y_2|\leqslant\cdots\leqslant|y_n|.$$

445

于是 $g_k(x)\leqslant g_k(y)$ 对所有 $k=1, 2, \cdots, n$ 成立的假定等价于 n 个不等式的组

$$\begin{aligned}
|x_n| &\leqslant |y_n|,\\
|x_{n-1}|+|x_n| &\leqslant |y_{n-1}|+|y_n|,\\
&\vdots \qquad\qquad\qquad\qquad (7.4.46)\\
|x_2|+\cdots+|x_n| &\leqslant |y_2|+\cdots+|y_n|,\\
|x_1|+|x_2|+\cdots+|x_n| &\leqslant |y_1|+|y_2|+\cdots+|y_n|. \qquad (*)
\end{aligned}$$

这些不等式与关于优化概念定义的不等式组(4.3.24)间的类似之处不仅仅是表面的.

如果这些不等式中的最后一个不等式(*)不是等式，通过缩小分量 y_1 的绝对值来修改 y 直到或者(a)不等式(*)是等式，或者(b) $|y_1|$ 缩小成零. 如果(b)出现在(a)之前，对下一个分量 y_2 重复这个步骤，如此做下去直到(a)出现. 其结果将产生一个修正向量 $y'=[y'_i]$，使得对 $i=1, \cdots, n$ 有 $|y'_i|\leqslant|y_i|$，对所有 $k=1, \cdots, n$，有 $g_k(x)\leqslant g_k(y')$，且(*)中等式成立. 由于绝对范数也是单调范数(5.5.10)，我们有 $g(y')\leqslant g(y)$. 因此，如果我们能证明 $g(x)\leqslant g(y)$ 对适合不等式组(7.4.46)(其中的(*)是等式)的任意 x，$y\in\mathbf{C}^n$ 成立，那么可以得知，$g(x)\leqslant g(y)$ 对适合一般的(7.4.46)的任意 x，$y\in\mathbf{C}^n$ 也成立.

假定(*)为等式的(7.4.46)成立就是假定向量 $-|x|=[-|x_i|]\in\mathbf{R}^n$ 优化向量 $-|y|=[-|y_i|]\in\mathbf{R}^n$(4.3.24)，而在这种情形，我们知道存在双随机矩阵 $S\in M_n$ 使得 $-|x|=S(-|y|)$ 或 $|x|=S|y|$(4.3.33). 因每个双随机矩阵是有限多个置换矩阵的凸组合(8.7.1)，我们可以把 S 写成 $S=\alpha_1P_1+\cdots+\alpha_NP_N$，其中，$\alpha_i\geqslant 0$，$\alpha_1+\cdots+\alpha_N=1$，且每个 $P_i\in M_n$ 是置换矩阵. 这时，有

$$\begin{aligned}
g(x) &= g(|x|)\\
&= g(S|y|) = g\Big(\sum_{i=1}^{N}\alpha_iP_i|y|\Big)\leqslant\sum_{i=1}^{N}g(\alpha_iP_i|y|)=\sum_{i=1}^{N}\alpha_ig(|y|)=g(|y|)\\
&= g(y),
\end{aligned}$$

这是因为 $g(\cdot)$ 是绝对向量范数，且 $g(\cdot)$ 是其自变量的分量的置换不变函数. □

定理的意义在于，为了使 $M_{m,n}$ 上的每个酉不变范数 $\|\cdot\|$ 有 $\|A\|\leqslant\|B\|$，必须而且只须

这个等式对樊畿 k 范数成立，$k=1, 2, \cdots, \min\{m, n\}$. [446]

7.4.47 推论 设 $A, B\in M_{m,n}$ 是分别具有奇异值 $\sigma_1(A)\geqslant\cdots\geqslant\sigma_q(A)\geqslant 0$ 和 $\sigma_1(B)\geqslant\cdots\geqslant\sigma_q(B)\geqslant 0$ 的某两个矩阵，其中 $q=\min\{m, n\}$. 要使 $\|A\|\leqslant\|B\|$ 对 $M_{m,n}$ 上的每个酉不变范数 $\|\cdot\|$ 成立，其充分条件是对所有 $i=1, 2, \cdots, q$ 有

$$\sigma_i(A)\leqslant\sigma_i(B), \tag{7.4.48}$$

而其必要充分条件是

$$\begin{aligned}\sigma_1(A)&\leqslant\sigma_1(B),\\ \sigma_1(A)+\sigma_2(A)&\leqslant\sigma_1(B)+\sigma_2(B),\\ &\vdots\\ \sigma_1(A)+\sigma_2(A)+\cdots+\sigma_q(A)&\leqslant\sigma_1(B)+\cdots+\sigma_q(B).\end{aligned}\tag{7.4.49}$$

证明： 所需要的关键论断是，$M_{m,n}$ 上的酉不变范数是其自变量的奇异值的对称度规函数(7.4.24). (7.4.48)的充分性只要求对称度规函数是单调范数的事实(5.5.10)，而关于不等式组(7.4.49)的更为明确的论断正是前一个定理的内容. □

为了应用推论(7.4.47)证明范数不等式，重述下面的事实常常是有用的：诸 Hermite 矩阵之和的有序特征值组成的向量优化各矩阵的有序特征值组成的向量之和.

7.4.50 引理 设 $A, B\in M_n$ 是具有有序特征值 $\lambda_1(A)\leqslant\cdots\leqslant\lambda_n(A)$ 和 $\lambda_1(B)\leqslant\cdots\leqslant\lambda_n(B)$ 的 Hermite 矩阵，又设 $\lambda_1(A-B)\leqslant\cdots\leqslant\lambda_n(A-B)$ 表示 $A-B$ 的有序特征值. 则向量

$$\lambda(A)-\lambda(B)=[\lambda_i(A)-\lambda_i(B)]$$

优化向量 $\lambda(A-B)=[\lambda_i(A-B)]$；即

$$\min\left\{\sum_{j=1}^{k}[\lambda_{i_j}(A)-\lambda_{i_j}(B)]:1\leqslant i_1<i_2<\cdots<i_k\leqslant n\right\}\geqslant\sum_{i=1}^{k}\lambda_i(A-B)$$

对 $k=1, 2, \cdots, n$ 成立，其中等式对 $k=n$ 成立.

证明： 定理(4.3.27)说明，由 $A-B+B=A$ 的特征值组成的向量 $\lambda(A)=\lambda((A-B)+B)=[\lambda_i((A-B)+B)]$ 优化向量 $\lambda(A-B)+\lambda(B)=[\lambda_i(A-B)+\lambda_i(B)]$，它等价于向量 $\lambda(A)-\lambda(B)$ 优化向量 $\lambda(A-B)$. □ [447]

以(7.4.47)中的条件及上述引理为工具，常常可以把关于范数或谱范数的逼近定理或不等式推广到整个酉不变范数类.

例如，(7.4.15)说明，如果 $A, B\in M_{m,n}$ 是分别具有有序奇异值 $\sigma_1(A)\geqslant\cdots\geqslant\sigma_q(A)\geqslant 0$ 和 $\sigma_1(B)\geqslant\cdots\geqslant\sigma_q(B)\geqslant 0$ 的某两个矩阵，且 $q=\min\{m, n\}$，则

$$\|A-B\|_2\geqslant\left(\sum_{i=1}^{q}[\sigma_i(A)-\sigma_i(B)]^2\right)^{1/2}$$

表示这个下界的另一种方式是

$$\|A-B\|_2\geqslant\|\Sigma(A)-\Sigma(B)\|_2,$$

其中，$A=V_1\Sigma(A)W_1^*$ 和 $B=V_2\Sigma(B)W_2^*$ 是奇异值分解，$\Sigma(A)$ 和 $\Sigma(B)$ 的“对角线”上的相应奇异值按从最大到最小的顺序排列. 这种不等式的另一个例子是(7.3.8(a))，它是关于谱范数的. (7.4.15)推广到所有酉不变范数就是取这种形式.

7.4.51 定理 设 A，$B\in M_{m,n}$ 是具有奇异值分解 $A=V_1\Sigma(A)W_1^*$ 和 $B=V_2\Sigma(B)W_2^*$ 的某两个矩阵，其中，V_1，$V_2\in M_m$ 和 W_1，$W_2\in M_n$ 是酉矩阵，而 $\Sigma(A)$ 和 $\Sigma(B)$ 的“对角元”都按递减顺序排列，则 $\|A-B\|\geqslant\|\Sigma(A)-\Sigma(B)\|$ 对 $M_{m,n}$ 上的每个酉不变范数 $\|\cdot\|$ 成立.

证明：设 $q=\min\{m,n\}$. 利用(7.3.7)把 A 的奇异值

$$\sigma_1(A)\geqslant\cdots\geqslant\sigma_q(A)\geqslant 0$$

与 Hermite 矩阵

$$\widetilde{A}=\begin{bmatrix}0 & A\\ A^* & 0\end{bmatrix}\in M_{m+n}$$

前 q 个非正特征值对应起来，$\widetilde{A}$ 的 $m+n$ 个有序特征值是

$$-\sigma_1(A)\leqslant -\sigma_2(A)\leqslant\cdots\leqslant -\sigma_q(A)\leqslant 0=\cdots=0\leqslant\sigma_q(A)\leqslant\cdots\leqslant\sigma_1(A),$$

对 $\widetilde{B}$ 和 $\widetilde{A}-\widetilde{B}$ 也可以作类似的对应. $\widetilde{A}$ 与 $\widetilde{B}$ 的有序特征值的差是 $\pm[\sigma_1(A)-\sigma_1(B)]$，…，$\pm[\sigma_q(A)-\sigma_q(B)]$以及($|m-n|$ 项)0. 虽然如何排出这个序列的顺序一般是不清楚的，但是，按照这个序列的顺序，q 个最小的元是$\{-|\sigma_i(A)-\sigma_i(B)|\}$，把引理(7.4.50)应用于 $\widetilde{A}$，

448 $\widetilde{B}$ 和 $\widetilde{A}-\widetilde{B}$ 便使我们确信

$$\sum_{i=1}^{k}-\sigma_i(A-B)\leqslant\min\left\{\sum_{j=1}^{k}-|\sigma_{i_j}(A)-\sigma_{i_j}(B)|:1\leqslant i_1<\cdots<i_k\leqslant n\right\}$$

对 $k=1$，…，q 成立，它等价于

$$\sum_{i=1}^{k}\sigma_i(A-B)\geqslant\max\left\{\sum_{j=1}^{k}|\sigma_{i_j}(A)-\sigma_{i_j}(B)|:1\leqslant i_1<\cdots<i_k\leqslant n\right\}$$

对 $k=1$，…，q 成立. 因为$\{|\sigma_i(A)-\sigma_i(B)|\}$是 $\Sigma(A)-\Sigma(B)$ 的奇异值的集合，推论(7.4.47)保证 $\|A-B\|\geqslant\|\Sigma(A)-\Sigma(B)\|$ 对任何酉不变范数 $\|\cdot\|$ 成立. □

7.4.52 例 定理(7.4.51)的一个推论是，对于在例(7.4.1)中所考虑的关于 Frobenius 范数求某个矩阵 $A\in M_n$ 的(在最小二乘意义下的)最佳秩 k 逼近问题作推广. 如果 $\|\cdot\|$ 是酉不变范数，又如果 $B\in M_n$ 有秩 k，则 $\sigma_1(B)\geqslant\cdots\geqslant\sigma_k(B)>0=\sigma_{k+1}(B)=\cdots=\sigma_n(B)$. 于是，

$$\begin{aligned}\|A-B\|&\geqslant\|\Sigma(A)-\Sigma(B)\|\\&=\|\mathrm{diag}(\sigma_1(A)-\sigma_1(B),\cdots,\sigma_k(A)-\sigma_k(B),\sigma_{k+1}(A),\cdots,\sigma_n(A))\|\\&\geqslant\|\mathrm{diag}(0,\cdots,0,\sigma_{k+1}(A),\cdots,\sigma_n(A))\|,\end{aligned}$$

其中，我们用到了对角矩阵的酉不变范数是单调范数的事实，这是因为它是诸对角元的对称度规函数. 另外，当 $B=VEW^*$ 时可能取等式，其中，$A=V\Sigma(A)W^*$ 是 A 的奇异值分解，而 $E=\mathrm{diag}[\sigma_1(A),\cdots,\sigma_k(A),0,\cdots,0]$.

因此，对任意 $A\in M_n$ 和秩为 k 的任意 $B\in M_n$，关于任意酉不变范数有下界

$$\begin{aligned}\|A-B\|&\geqslant\|\mathrm{diag}(0,\cdots,0,\sigma_{k+1}(A),\cdots,\sigma_n(A))\|\\&\geqslant\sigma_n(A)\|\mathrm{diag}(0,\cdots,0,1,\cdots,1)\|\end{aligned}$$

(最后一个表示式的对角线上有 k 项零)，其中，第一个不等式可以是等式，而第二个不等式一般不取等式. 第二个不等式(如果 A 是非奇异矩阵，该不等式完全可以从对称度规函数的单调性推出，而如果 A 是奇异矩阵，则结论是明显的)有以下优点：它对范数的依赖关系只是 k 的

函数而不是 A 的函数. 特别是, 这说明, 对任意非奇异矩阵 $A\in M_n$ 和任意酉不变范数 $\|\cdot\|$ 有最大下界

$$\|A-B\|\geqslant\sigma_n(A)\|\operatorname{diag}(0,\cdots,0,1)\|, \tag{7.4.53}$$

449

它对 A 与任意奇异矩阵 B 间的距离成立: 即 A 到奇异矩阵所组成的闭集的最小距离(关于酉不变范数 $\|\cdot\|$)是 $\sigma_n(A)\|\operatorname{diag}(0,\cdots,0,1)\|$.

7.4.54 例 我们可以利用对称度规函数的性质给出 M_n 上的酉不变范数是矩阵范数的简单特征. 如果 $\|\cdot\|$ 是 M_n 上的酉不变矩阵范数, 则从推论(5.6.35)得知, $\|A\|\geqslant\sigma_1(A)$ 对所有 $A\in M_n$ 成立. 利用定理(5.6.9)和 M_n 上的每个酉不变范数是自伴范数的事实(见习题 2), 用论断 $[\sigma_1(A)]^2=\rho(A^*A)\leqslant\|A^*A\|\leqslant\|A^*\|\|A\|=\|A\|^2$ 也可以直接证明上述结论. 另一方面, 设 $\|\cdot\|$ 是使 $\|A\|\geqslant\sigma_1(A)$ 对所有 $A\in M_n$ 都成立的 M_n 上的酉不变范数, 又设 g 是由 $\|\cdot\|$ 诱导的 $\mathbf{C}^n$ 上的对称度规函数. 利用 7.3 节习题 18 中给出的类似于 Weyl 不等式的关于奇异值的乘法不等式以及 g 是单调范数的事实便可推出

$$\begin{aligned}\|AB\|&=g(\sigma_1(AB),\sigma_2(AB),\cdots,\sigma_n(AB))\\&\leqslant g(\sigma_1(A)\sigma_1(B),\sigma_1(A)\sigma_2(B),\cdots,\sigma_1(A)\sigma_n(B))\\&=\sigma_1(A)g(\sigma_1(B),\sigma_2(B),\cdots,\sigma_n(B))\\&=\sigma_1(A)\|B\|\leqslant\|A\|\|B\|.\end{aligned}$$

因此, M_n 上的酉不变范数 $\|\cdot\|$ 是矩阵范数, 当且仅当 $\|A\|\geqslant\sigma_1(A)=\|A\|_2$ 对所有 $A\in M_n$ 成立. 特别是, 所有樊畿 k 范数, $k=1, 2, \cdots, n$, 以及所有 Schatten p 范数, $p\geqslant1$, [它们分别是由(7.4.4)中的对称度规函数及(5.2.4)诱导的]是矩阵范数. 这个特征的另一个推论是, M_n 上的酉不变矩阵范数所组成的集合是凸集, M_n 上的所有矩阵范数所组成的集合不是凸集[参看(5.6)节习题 9].

习题

1. 设 $A\in M_{m,n}$ 的秩 $k>0$. 假定求一个秩为 $k_1<k$ 的矩阵 $A_1\in M_{m,n}$ 要求能按 Frobenius 范数最佳逼近 A. 说明这可以按下述方式进行:

设 $A=V\Sigma W^*$ 是 A 的奇异值分解. 设 Σ_1 除了仅取 $\sigma_1, \cdots, \sigma_{k_1}$ 而其所余下 $n-k_1$ 个"对角"元为零以外, Σ_1 与 Σ 是相同的. 于是 $A_1\equiv V\Sigma_1W^*$ 有所要求的性质. **提示:** 利用(7.4.15). 注意到(7.4.52)证明了所给逼近不仅关于 Frobenius 范数是"最佳"的, 而且关于所有酉不变范数也是"最佳"的.

2. M_n 上的范数称为自伴范数, 是指 $\|A\|=\|A^*\|$ 对每个 $A\in M_n$ 成立. 试用定理(7.4.24)证明, M_n 上的每个酉不变范数是自伴范数. 试给一个是自伴范数而不是酉不变范数的例子.

450

3. 试用定理(7.4.10)和例(7.4.6)的方法来确定, 用一个具有标准正交行的矩阵 $Y\in M_{m,n}$ 的纯量倍对一个给定矩阵 $A\in M_{m,n}$(其中 $m\leqslant n$)的最佳最小二乘逼近. **提示:** 证明, 这样的矩阵 Y 一定有形式 $Y=VDW$, 其中 $V\in M_m$ 和 $W\in M_n$ 是酉矩阵, $D=[I\ 0]\in M_{m,n}$, $I\in M_m$, 而 $0\in M_{m,n-m}$. 极小化 $\|A-cY\|_2^2$ 的问题与极小化 $\|A\|_2^2-(\operatorname{Re}\operatorname{tr}AY^*)^2/m$ 是相同的. 如果

$A=V_1\Sigma W_1^*$ 是 A 的奇异值分解，说明这个极小化问题变为求

$$\max \operatorname{Re} \operatorname{tr}\{\Sigma WD^*V: W \in M_n \text{ 和 } V \in M_n \text{ 是酉矩阵}\},$$

然后利用定理(7.4.10)解这个如同例(7.4.13)中的问题. 说明这种情形下的误差值与例(7.4.6)有相同的形式.

4. 考虑对角矩阵 A，$B\in M_n$，证明所有可能的排列 τ 可以出现在(7.4.11)中.

5. 考虑(7.4.7)中定义的函数 $u(A)$. 证明

$$u(A) \leqslant \sqrt{n}\,\|A\|_2$$

对所有 $A\in M_n$ 成立，且这个界是可达到的. 试用定义直接证明 $u(A)$是 M_n 上的向量范数，并且说明为什么 $u(A)$实际上是 M_n 上的矩阵范数. **提示**：参看例(7.4.54).

6. 证明，如果 $A\in M_n$ 是非奇异矩阵，且 $\kappa(A)=\|A\|_2\|A^{-1}\|_2$ 是 A 关于谱范数的条件数，则 $\kappa(A)=\sigma_1/\sigma_n$，最大奇异值和最小奇异值之比. 这如何同估计 $\kappa(A)\geqslant|\lambda_1/\lambda_n|$ 作比较?

7. 证明 Kantorovich 不等式(7.4.42)中的常量是 λ_1 和 λ_n 的几何平均值与 λ_1 和 λ_n 的算术平均值之比的平方.

8. 设 $A\in M_n$ 是非奇异 Hermite 矩阵. 试用 Kantorovich 不等式(7.4.40)证明

$$\max_{x\neq 0} \frac{\|Ax\|_2\|A^{-1}x\|_2}{\|x\|_2^2}=\frac{\sigma_1^2+\sigma_n^2}{2\sigma_1\sigma_n}=\frac{1}{2}\left(\frac{\sigma_1}{\sigma_n}+\frac{\sigma_n}{\sigma_1}\right),$$

其中 $\sigma_1\geqslant\cdots\geqslant\sigma_n>0$ 是 A 的奇异值. 证明 σ_1 和 σ_n 分别是 A 的诸特征值的最大绝对值和最小绝对值，并且证明

151

$$\frac{1}{2}\left(\frac{\sigma_1}{\sigma_n}+\frac{\sigma_n}{\sigma_1}\right)=\frac{1}{2}(\kappa+\kappa^{-1}),$$

其中 κ 是 A 的谱条件数. 给出一个向量 x 使上述极大值可达到. 利用谱条件数以及它与上面所定义的极大值的关系，说明为什么有

$$\frac{1}{2}\left(\frac{\sigma_1}{\sigma_n}+\frac{\sigma_n}{\sigma_1}\right)\leqslant\frac{\sigma_1}{\sigma_n}$$

直接证明这个不等式. **提示**：证明对于 $x\geqslant1$，$f(x)=x-[x+(1/x)]/2$ 是增函数.

9. 设 λ_1，λ_2，…，λ_n 是 n 个给定的正实数. 试用 Kantorovich 不等式(7.4.42)证明，如果 α_1，…，α_n 是非负的，且其和为 1，则

$$\left(\sum_{i=1}^{n}\alpha_i\lambda_i\right)\left(\sum_{i=1}^{n}\frac{\alpha_i}{\lambda_i}\right)\leqslant\frac{(\lambda_{\max}+\lambda_{\min})^2}{4\lambda_{\max}\lambda_{\min}}.$$

10. 证明(属于 Greub 和 Rheinboldt 的)Kantorovich 不等式(7.4.42)的下述推广：设 B，$C\in M_n$ 是交换的正定矩阵，分别有特征值 $0<\lambda_1\leqslant\cdots\leqslant\lambda_n$ 和 $0<\mu_1\leqslant\cdots\leqslant\mu_n$，则

$$(x^*BCx)^2\geqslant\frac{4\lambda_1\lambda_n\mu_1\mu_n}{(\lambda_1\mu_1+\lambda_n\mu_n)^2}(x^*B^2x)(x^*C^2x)$$

对所有 $x\in\mathbf{C}^n$ 都成立. **提示**：因为对某个酉矩阵 $U\in M_n$ 有 $B=U\Lambda U^*$ 和 $C=UMU^*$，先用 $y=U^*x$ 写所要求的不等式，然后用 $z=(\Lambda M)^{1/2}y$ 写不等式. 于是代 $B=\Lambda M$ 应用(7.4.41)可证明所要求的不等式成立(且可取等式)，并且对于指标 $1\leqslant j\neq k\leqslant n$ 的某个选择有形如

$$\frac{\lambda_1\lambda_n\mu_j\mu_k}{(\lambda_1\mu_j+\lambda_n\mu_k)^2}$$

的常数．证明这个形式的最小常数当 $j=1$ 和 $k=n$ 时出现．但是，这最后的推广不等式不可能取等号．

11．简化 Kantorovich 不等式(7.4.42)，证明，若 $B\in M_n$ 是正定矩阵，则对所有 $x\in \mathbf{C}^n$ 有

$$(x^*Bx)(x^*B^{-1}x)\geqslant \|x\|_2^4.$$

更一般地，若 $B\in M_n$ 是正定矩阵，证明，则对所有 x，$y\in \mathbf{C}^n$ 有

$$(x^*Bx)(y^*B^{-1}y)\geqslant (x^*y)^2,$$

其中等式对 $x=B^{-1}y$ 成立．由此得出对所有 $x\in \mathbf{C}^n$ 有

$$(x^*x)^2\leqslant (x^*Bx)(x^*B^{-1}x)\leqslant \frac{[(\lambda_1+\lambda_n)/2]^2}{\lambda_1\lambda_n}(x^*x)^2.$$

452

提示：如果 $\lambda_i>0$，证明

$$\left|\sum_{i=1}^n x_i\bar{y}_i\right|^2=\left|\sum_{i=1}^n(\sqrt{\lambda_i}x_i)\left(\frac{\bar{y}_i}{\sqrt{\lambda_i}}\right)\right|^2\leqslant\left(\sum_{i=1}^n\lambda_i|x_i|^2\right)\left(\sum_{i=1}^n\frac{|y_i|^2}{\lambda_i}\right).$$

然后记 $B=U\Lambda U^*$．

12．设 $B\in M_n$ 是正定矩阵，$y\in \mathbf{C}^n$ 是任一非零向量，且定义

$$f(B,y)\equiv\min\left\{\frac{x^*Bx}{(x^*y)^2}:x\in \mathbf{C}^n,\quad x^*y\neq 0\right\}.$$

证明 $f(B,y)$ 是有意义的，然后利用习题 11 证明 $f(B,y)=1/y^*B^{-1}y$．证明 f 具有超加性性质，即对所有 $y\in \mathbf{C}^n$ 和所有正定矩阵 A，$B\in M_n$ 有

$$f(A+B,y)\geqslant f(A,y)+f(B,y)$$

现在设 $y=e_i$，e_i 是第 i 个标准单位基向量，然后推出 Bergstrom 不等式

$$\frac{\det(A+B)}{\det(A_i+B_i)}\geqslant\frac{\det A}{\det A_i}+\frac{\det B}{\det B_i},\quad i=1,\cdots,n$$

对任何正定矩阵 A，$B\in M_n$ 成立，其中 $A_i\in M_{n-1}$ 表示划去 A 的第 i 行和第 i 列得到的 A 的主子矩阵，B_i 的意义类似．这种接近于 Bergstrom 不等式的方法是有着广泛用途的所谓拟线性化方法的一个应用实例；拟线性化是把一个所考虑的量的非线性函数表示成另一个函数的约束极值，而这个新函数线性地(或许只是加性地)依赖于所考虑的量．在(7.4.24)中的关键步骤(证明用奇异值的对称度规函数定义的 $M_{m,n}$ 上的准范数实际是一个范数)是用(5.4.12)中的拟线性化完成的．

13．对于任一复数 z，不等式 $|z-\mathrm{Re}\,z|\leqslant|z-x|$ 对任何实数 x 成立．这个不等式到方阵 $A\in M_n$ 的看似合理的推广是

$$\left\|A-\frac{1}{2}(A+A^*)\right\|\leqslant\|A-H\|$$

对所有 Hermite 矩阵 $H\in M_n$ 成立．证明，这个不等式对所有酉不变范数 $\|\cdot\|$ 以及更一般地对所有自伴范数成立．由此得出，(关于 $\|\cdot\|$)从一个给定矩阵 $A\in M_n$ 到由 M_n 中的 Hermite 矩阵组成的闭集的距离是 $\frac{1}{2}\|A-A^*\|$．**提示**：$A-\frac{1}{2}(A+A^*)=\frac{1}{2}(A-H)+\frac{1}{2}(H-A^*)$，因

[453] 而$\left\|A-\frac{1}{2}(A+A^*)\right\|\leqslant\frac{1}{2}\|A-H\|+\frac{1}{2}\|H-A^*\|$.

14. 对任意复数z，有不等式 $|\mathrm{Re}\, z|\leqslant|z|$. 证明它的明显推广$\|(A+A^*)/2\|\leqslant\|A\|$对所有$A\in M_n$和所有酉不变(甚至自伴)范数$\|\cdot\|$成立.

15. 设$A\in M_n$是给定的，$\lambda_1\leqslant\cdots\leqslant\lambda_n$是$\frac{1}{2}(A+A^*)$的有序特征值，又设$\sigma_1\geqslant\cdots\geqslant\sigma_n$是$A$的有序奇异值. 说明为什么不等式

$$\lambda_{n-k+1}\left(\frac{1}{2}[A+A^*]\right)\leqslant\sigma_k(A),\quad k=1,\cdots,n$$

可以看作关于复数的不等式$\mathrm{Re}\, z\leqslant|z|$的一个推广. 这个不等式是说，$A$的第$k$个最大奇异值大于或等于$\frac{1}{2}(A+A^*)$的第$k$个最大特征值. **提示**：若$y$是Euclid单位向量，则

$$\frac{1}{2}y^*(A+A^*)y=\mathrm{Re}\, y^*Ay\leqslant\|Ay\|_2.$$

用Courant Fischer定理(4.2.11)表示λ_{n-k+1}，然后用这个不等式和(7.3.10)得到σ_k.

16. 设$A\in M_n$是给定的，又设$\|\cdot\|$是M_n上的一个酉不变范数. 利用(7.4.51)证明，$\|A-U\|\geqslant\|\Sigma(A)-I\|$对任一酉矩阵$U\in M_n$成立，且这个不等式可取等式. 由此得出$\|\Sigma(A)-I\|$是(关于$\|\cdot\|$)从$A$到由$M_n$中的酉矩阵组成的紧集的距离.

17. 设$A\in M_n$有奇异值分解$A=V\Sigma(A)W^*$，又设$\|\cdot\|$是M_n上的酉不变范数. 证明

$$\|\Sigma(A)-I\|\leqslant\|A-U\|\leqslant\|\Sigma(A)+I\|$$

对任何酉矩阵$U\in M_n$成立. **提示**：证明，在任一酉矩阵的任一奇异值分解中有$\Sigma(U)=I$，于是从(7.4.51)直接推出下界成立. 关于上界，利用(7.3)节16题中类似于Weyl的加法特征值不等式的奇异值不等式证明$\sigma_{i+j-1}(A+(-U))\leqslant\sigma_i(A)+\sigma_j(-U)$. 然后用(7.4.48).

18. 试以例(7.4.53)中关于非奇异矩阵A的不等式为指南，求$\|A-B\|$的最大下界，其中，$A\in M_n$是给定的秩k_1矩阵，$B\in M_n$是任意秩$k<k_1$矩阵，而$\|\cdot\|$是酉不变范数.

进一步阅读 定理(7.4.21)$m=n$的情形的最初原型属于Von Neumann；可参看(5.4)节引用的文章. Wielandt和Kantorovich不等式取自[Hou 64]，并且作了改编，[Hou 64]还有[454]许多原始资料，有关的推广以及其他资料可参看A. Clausing, "Kantorovich-Type Inequalities," *Amer. Math. Monthly* 89(1982), 314-320. 习题12中接近于Bergstrom不等式的方法取自[BB]，它用一大章(还有大量的参考资料)专门论述由正定矩阵引起的诸不等式；并且对拟线性化方法也有讨论，还给出了许多例子. 关于对所有酉不变范数都成立的诸不等式的其他资料可参看L. Mirsky, "Symmetric Gauge Functions and Unitarily Invariant Norms," *Quart. J. Math. Oxford* 11(2)(1960), 50-59以及K. Fan and A. J. Hoffman, "Some Metric Inequalities in the Space of Matrices," *Proc. Amer. Math. Soc.* 6(1955), 111-116. 例如，这些结果如何应用于统计学以及有关统计学文献的其他资料可参看C. R. Rao, "Matrix Approximations and Reduction of Dimensionality in Multivariate Statistical Analysis," *Multivariate Analysis V*, Proceedings of the Fifth International Symposium on Multivariate Analysis, P. R. Krishnaiah, North-Holland, Amsterdam, 1980, pp. 1-22.

7.5 Schur 乘积定理

一个特别简单(又形式自然)的矩阵"乘法"是按对应分量相乘.

7.5.1 定义 如果 $A=[a_{ij}]\in M_{m,n}$ 和 $B=[b_{ij}]\in M_{m,n}$ 是给定的矩阵，则 A 和 B 的 Hadamard 乘积是矩阵 $A\circ B=[a_{ij}b_{ij}]\in M_{m,n}$.

常常称 Hadamard 乘积为 Schur 乘积. 像矩阵加法一样，Hadamard 乘法是可交换的，因而它比普通矩阵乘法要简单得多.

Hadamard 乘积自然可以从几种不同的观点产生出来. 例如，$f(\theta)$ 和 $g(\theta)$ 是周期为 2π 的连续周期函数，又如果

$$a_k=\int_0^{2\pi}e^{ik\theta}f(\theta)\mathrm{d}\theta \quad 和 \quad b_k=\int_0^{2\pi}e^{ik\theta}g(\theta)\mathrm{d}\theta,$$

$k=0,\ \pm1,\ \pm2,\ \cdots$，则卷积

$$h(\theta)=\int_0^{2\pi}f(\theta-t)g(t)\mathrm{d}t$$

有三角矩

$$c_k=\int_0^{2\pi}e^{ik\theta}h(\theta)\mathrm{d}\theta,$$

455

它适合恒等式 $c_k=a_kb_k$，$k=0,\ \pm1,\ \pm2,\ \cdots$. 于是，$h(\theta)$ 的三角矩的 Toeplitz 矩阵是 $f(\theta)$ 和 $g(\theta)$ 的三角矩的 Toeplitz 矩阵的 Hadamard 乘积：

$$[c_{i-j}]=[a_{i-j}]\circ[b_{i-j}].$$

如果 $f(\theta)$ 和 $g(\theta)$ 都是非负实值函数，则卷积 $h(\theta)$ 也是非负实值函数. 因此，如(7.0.5)中所证明的那样，矩阵 $[a_{i-j}]$, $[b_{i-j}]$ 以及 $[c_{i-j}]$ 都是半正定矩阵. 这是 Schur 乘积定理的一个实例，该定理是说：两个半正定矩阵的 Hadamard 乘积是半正定矩阵.

又如，考虑积分算子

$$K(f)=\int_a^bK(x,y)f(y)\mathrm{d}y,$$

其中，核 $K(x,y)$ 是有限区间 $[a,b]\times[a,b]$ 上的连续函数，且 $f\in C[a,b]$. 如果我们有第二个核 $H(x,y)$，则可以考虑(逐点)乘积核 $L(x,y)=K(x,y)H(x,y)$ 和相关联的积分算子

$$L(f)\equiv\int_a^bL(x,y)f(y)\mathrm{d}y=\int_a^bK(x,y)H(x,y)f(y)\mathrm{d}y.$$

线性映射 $f\to K(f)$ 自然是矩阵-向量乘法的极限(把积分看作有限 Riemann 和的逼近)，并且积分算子的许多性质可以通过对矩阵的已知结果取适当的极限导出. 从这个观点出发，积分核的(逐点)乘积导出一个积分算子，它是类似于矩阵的 Hadamard 乘积的自然的连续形式.

如果积分核 $K(x,y)$ 有以下性质，

$$\int_a^b\int_a^bK(x,y)f(x)\bar{f}(y)\mathrm{d}x\mathrm{d}y\geqslant0$$

对所有 $f\in C[a,b]$ 成立，则称 $K(x,y)$ 为*半正定核*. 一个经典结果(Mercer 定理)是说，如果

$K(x, y)$是有限区间$[a, b]$上的连续半正定核，则存在一组正实数$\{\lambda_i\}$(称为"特征值")和一组连续函数$\{\phi_i(x)\}$(称为"特征函数")使得在$[a, b]\times[a, b]$上有

$$K(x,y) = \sum_{i=1}^{\infty} \frac{\phi_i(x)\overline{\phi}_i(y)}{\lambda_i}.$$

[456] 且该级数绝对一致收敛.

如果$K(x, y)$和$H(x, y)$是同一个有限区间$[a, b]$上的连续半正定核，则$H(x, y)$在$[a, b]\times[a, b]$上也有绝对一致收敛的表达式

$$H(x,y) = \sum_{i=1}^{\infty} \frac{\psi_i(x)\overline{\psi}_i(y)}{\mu_i},$$

其中，所有的$\mu_i>0$. 根据相应级数的直接乘法，(逐点)乘积核$L(x, y)=K(x, y)H(x, y)$在$[a, b]\times[a, b]$上有表达式

$$L(x,y) = \sum_{i,j=1}^{\infty} \frac{\phi_i(x)\psi_j(x)\overline{\phi}_i(y)\overline{\psi}_j(y)}{\lambda_i\mu_j},$$

它也绝对一致收敛. 于是

$$\int_a^b\int_a^b L(x,y)f(x)\overline{f}(y)\mathrm{d}x\mathrm{d}y = \sum_{i,j=1}^{\infty} \frac{1}{\lambda_i\mu_j}\left|\int_a^b \phi_i(x)\psi_j(x)f(x)\mathrm{d}x\right|^2 \geqslant 0,$$

因而$L(x, y)$也是半正定的. 这是 Schur 乘积定理的另一个实例.

练习 证明，两个 Hermite 矩阵的 Hadamard 乘积总是 Hermite 矩阵，但是，两个 Hermite 矩阵的普通矩阵乘积是 Hermite 矩阵，当且仅当它们是可交换的.

练习 考察矩阵$A=\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}$和$B=\begin{bmatrix}2 & 1\\ 1 & 3\end{bmatrix}$. 证明，$A$，$B$和$A\circ B$是半正定矩阵，而普通的矩阵乘积$AB$不是半正定矩阵. 但是可以证明$AB$的特征值是正的.

在这里引进 Hadamard 乘积的主要理由是，(不同于普通矩阵乘积)它使由诸半正定矩阵组成的锥不变，它还给出了半正定矩阵和非负实数间的另一种类似. 我们从一个具有独立意义的论断开始.

任意矩阵可以表示成一些秩 1 矩阵的和，其中被加矩阵的个数等于该矩阵的秩，而对于半正定矩阵，被加矩阵也可以选为半正定矩阵.

7.5.2 定理 如果$A\in M_n$是秩k半正定矩阵，则A可写成形式

[457]
$$A = v_1v_1^* + v_2v_2^* + \cdots + v_kv_k^*,$$

其中，每个$v_i\in \mathbf{C}^n$，且$\{v_1, \cdots, v_k\}$是非零向量的正交组.

证明：利用谱定理将A写成$A=U\Lambda U^*$，且设v_i是U的第i列的$\lambda_i^{1/2}$倍. □

我们的主要结果常常称为 Schur 乘积定理.

7.5.3 定理 如果A，$B\in M_n$是半正定矩阵，则$A\circ B$也是半正定矩阵. 此外，如果A和B都是正定矩阵，则$A\circ B$也是正定矩阵.

证明：利用(7.5.2)记$A=v_1v_1^*+\cdots+v_kv_k^*$和$B=w_1w_1^*+\cdots+w_mw_m^*$，其中$k=\operatorname{rank} A$且$m=\operatorname{rank} B$. 注意到

$$A \circ B = \sum_{i,j=1}^{k,m} u_{ij} u_{ij}^*,$$

其中 $u_{ij} = v_i \circ w_j$. 因为 $A \circ B$ 是(秩 1)半正定矩阵之和, 所以 $A \circ B$ 也是半正定矩阵.

如果 A 和 B 都是正定矩阵, 则 $k=m=n$, 且向量组 $\{v_i\}$ 和 $\{w_i\}$ 都是 $\mathbf{C}^n$ 的正交基. 尚若 $A \circ B$是奇异矩阵, 则存在某个非零向量 x 使得$(A \circ B)x=0$, 因而

$$x^*(A \circ B)x = \sum_{i,j=1}^{k,m} x^*(u_{ij}u_{ij}^*)x = \sum_{i,j=1}^{k,m} |x^* u_{ij}|^2 = 0.$$

另一方面, 每一项必须分别为零, 因而

$$|x^* u_{ij}|^2 = |x^*(v_i \circ w_j)|^2 = |(x \circ \bar{v}_i)^* w_j|^2 = 0$$

对所有 i 和 j 成立. 这说明, 对每个 i, 向量 $x \circ v_i$ 正交于所有向量 $w_1, w_2, \cdots, w_n$, 因此, $x \circ \bar{v}_i = 0$ 对所有 $i=1, 2, \cdots, n$ 成立. 特别是, 这蕴涵 $v_i^* x = 0$ 对所有 $i=1, \cdots$成立. 因为这意味着 x 与基的所有元素正交, 所以一定有 $x=0$. 由此得出 $A \circ B$ 一定是非奇异矩阵. □

练习 设 $A, B \in M_n$. 试用(7.5.3)的证法证明, 当 A 和 B 是半正定矩阵时, 总有 $\operatorname{rank} A \circ B \leqslant (\operatorname{rank} A)(\operatorname{rank} B)$. 特别是证明, 如果$(\operatorname{rank} A)(\operatorname{rank} B) < n$, 则 $A \circ B$ 一定是奇异矩阵.

练习 证明, 当 A 和 B 是 Hermite 矩阵而不一定是半正定矩阵时, 上述练习的论断仍然成立.

458

练习 考察矩阵 $A=\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$和 $B=\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, 说明, 即使 A 和 B 都有正的秩, $\operatorname{rank} A \circ B$ 也可能是零.

练习 证明, 如果 A 是正定矩阵而 B 是负定矩阵, 则 $A \circ B$ 是负定矩阵.

7.5.4 推论(Fejer 定理) 设 $A=[a_{ij}] \in M_n$, 则 A 是半正定矩阵当且仅当

$$\sum_{i,j=1}^{n} a_{ij} b_{ij} \geqslant 0$$

对所有半正定矩阵 $B=[b_{ij}] \in M_n$ 成立.

证明: 假定 A 和 B 是半正定矩阵, 且设 $x \in \mathbf{C}^n$ 是所有分量都等于 1 的向量. 于是 $A \circ B$ 是半正定矩阵, 而且 $\Sigma a_{ij} b_{ij}$ 正好是 $x^*(A \circ B)x$, 它肯定是非负的. 反过来, 如果当 B 是半正定矩阵时有 $\Sigma a_{ij} b_{ij} \geqslant 0$, 此时对任意给定的向量 $x \in \mathbf{C}^n$ 令 $B=[b_{ij}] \equiv [\bar{x}_i x_j]$. 于是 B 是半正定矩阵, 且

$$\sum_{i,j=1}^{n} a_{ij} b_{ij} = \sum_{i,j=1}^{n} a_{ij} \bar{x}_i x_j = x^* A x \geqslant 0.$$

因为 $x \in \mathbf{C}^n$ 是任意的, 由此得出 A 是半正定矩阵. □

7.5.5 应用 设 $D \subset \mathbf{R}^n$ 是有界开集. 由

$$Lu \equiv \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x) \frac{\partial u}{\partial x_i} + c(x)u \tag{7.5.6}$$

给出的 $C^2(D)$上的二阶线性微分算子 L 称为在 D 中是椭圆型的, 是指对所有的 $x \in D$, 矩阵 $A(x) \equiv [a_{ij}(x)]$是正定矩阵. 假定存在某个 $u \in C^2(D)$, 它在 D 的闭包上连续且在 D 中适合方

程 $Lu=0$. 关于函数 u 在 D 中的局部极大和局部极小，我们能说些什么？假定 $y\in D$ 是关于 u 的局部极小值点．于是对所有 $i=1, 2, \cdots, n$ 有

$$\left.\frac{\partial u}{\partial x_i}\right|_y=0,$$

并且 Hessian 矩阵

459

$$\left[\frac{\partial^2 u}{\partial x_i \partial x_j}\right]$$

在点 y 是正定的．因此，在点 y 有

$$Lu=0=\sum_{i,j=1}^{n} a_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j}+cu,$$

并且由 Fejer 定理(7.5.4)可知，含有二阶导数的和必须是非负的，因而项 $c(y)u(y)$ 必须是非正的．特别是，如果 $c(y)<0$，则不可能有 $u(y)<0$．同理可证，如果 $c(y)<0$，$u(y)$ 不可能在相对极大值内点 y 为正．这些简单的结论就是下述重要原理的要点．

7.5.7 弱极小原理 设用(7.5.6)定义的算子 L 为 D 中的椭圆型算子，且假定在 D 中 $c(x)<0$．如果 $u\in C^2(D)$ 在 D 中适合 $Lu\equiv0$，则 u 不能有负的相对极小值内点，或者不可能有正的相对极大值点．此外，如果 u 在 D 的闭包上连续且在 D 的边界上非负，则 u 在 D 中必定处处非负．

从极小原理可以推出偏微分方程的一个基本的唯一性定理：

7.5.8 Fejer 唯一性定理 假定用(7.5.6)定义的算子 L 是椭圆型的，设在 D 中 $c(x)<0$，并且考虑下述边值问题：

在 D 中，$Lu=f$，f 是已知函数；

在 ∂D 上，$u=g$，g 是已知函数；

在 D 中，u 二次连续可微；

在 D 的闭包上 u 连续．

则这个问题至多存在一个解．

证明： 假如 u_1 和 u_2 是这个问题的两个解，则函数 $v=u_1-u_2$ 是同一类型问题的一个解，不过具有零边值条件且在 D 中 $Lv=0$．由弱极小原理可知，在 D 中 v 必须非负．把同样的论证用于 $-v$，得知 v 在 D 中也必须非正，因而在 D 中 $v\equiv0$． □

练习 说明弱极小原理和 Fejer 唯一性定理如何应用于 $D\subset\mathbf{R}^n$ 中的偏微分方程 $\nabla^2 u-\lambda u=0$，其中 λ 是正实参数．

460

Schur 乘积定理的最后一个推论是容易证明的．如果 $A=[a_{ij}]\in M_n$ 是半正定矩阵，那么 $A\circ A=[a_{ij}^2]$ 也是半正定矩阵．由归纳法可知，对所有 $k=1, 2, \cdots$，所有正整数次 Hadanard 幂 $[a_{ij}^k]$ 是半正定矩阵．因为半正定矩阵的任意非负线性组合是半正定矩阵(7.1.3)，这蕴涵，只要所有 $a_i\geqslant0$，

$$a_0 J+a_1 A+a_2 A\circ A+\cdots+a_m \overbrace{A\circ\cdots\circ A}^{m\text{项}}=[a_0+a_1 a_{ij}+a_2 a_{ij}^2+\cdots+a_m a_{ij}^m]$$
$$=[p(a_{ij})]$$

就是半正定矩阵；$p(x)=a_0+a_1x+\cdots+a_m x^m$ 是非负系数多项式．更一般地，如果

$$f(z) = \sum_{k=0}^{\infty} a_k z^k$$

是解析函数，其中所有 $a_k \geq 0$ 且收敛半径 $R>0$，则经简单的极限证法可证，只要所有 $|a_{ij}| < R$，$[f(a_{ij})] \in M_n$ 就是半正定矩阵. 或许最简单的例子是 $f(z)=e^z$，其幂级数对所有 $z \in \mathbf{C}$ 都收敛，它的系数是 $a_k = 1/k! > 0$. 根据这番证明，只要 $A=[a_{ij}] \in M_n$ 是半正定矩阵，$[e^{a_{ij}}]$ 就是半正定矩阵. 这个结果还可以改进；A 的较弱条件足以保证以 A 的分量为指数的矩阵是半正定的. 见[HJ].

7.5.9 推论 设 $A=[a_{ij}] \in M_n$ 是半正定矩阵. 则

(a) 对所有 $k=1, 2, \cdots$，矩阵$[a_{ij}^k]$是半正定矩阵.

(b) 如果 $f(z)=a_0+a_1 z+a_2 z^2+\cdots$是具有非负系数的解析函数，且收敛半径 $R>0$，则当所有 $|a_{ij}|<R$ 时，矩阵$[f(a_{ij})]$是半正定矩阵.

习题

1. 证明，如果 $H(A)$(A 的 Hermite 部分)是正定矩阵，且 B 是正定矩阵，则 $H(A \circ B)$是正定矩阵.

2. 如果 $A=[a_{ij}]$是半正定矩阵，证明矩阵$[|a_{ij}|^2]$也是半正定矩阵. **提示：**考虑 $A \circ \bar{A}$.

3. 如果 $A=[a_{ij}] \in M_n$ 是半正定矩阵，证明对所有 $\lambda \in \mathbf{R}$，矩阵$[e^{e^{a_{ij}}\lambda}]$是半正定的. 461

4. 如果 $A=[a_{ij}] \in M_n$ 是半正定矩阵，则正整数次 Hadamard 幂矩阵 $A^{(k)}$ 和 Hadamard 绝对值平方矩阵 $A \circ \bar{A}$ 恒为半正定矩阵. 但是，关于 Hadamard 绝对值矩阵 $|A| \equiv [|a_{ij}|]$，其结果又如何呢？(a)假定 $A \in M_n$ 是正定矩阵. 对 $n=1, 2, 3$，利用行列式准则(7.2.5)直接证明 $|A|$ 是正定矩阵. 试用取极限的办法对半正定矩阵 A(仅当 $n=1, 2, 3$ 时)得到相应的结果. (b)利用 $f(x)\cos(x)$是正定函数的事实[或把 $\cos(x)$写成 $\cos(x)=(e^{ix}+e^{-ix})/2$ 然后计算二次型]证明，对$\{x_1, x_2, \cdots, x_n\} \subset \mathbf{R}$ 的所有选择和所有 $n=1, 2, \cdots$，矩阵 $A=[\cos(x_i-x_j)]$是半正定矩阵. (c)设 $n=4$，且设 $x_1=0$，$x_2=\pi/4$，$x_3=\pi/2$，和 $x_4=3\pi/4$. 在这种情形，直接计算(b)中的(一定是半正定的)矩阵 A，并且得出它是 Toenitz 矩阵. 计算 $|A|$ 和 $\det|A|$，且证明 $|A|$ 不可能是半正定矩阵.

5. 考虑习题 4 中的矩阵 $|A|$，证明，$B \equiv |A| \circ |A|$ 是半正定矩阵，而它的非负"Hadamard 方根"不是半正定矩阵. 将这与普通方根 $B^{1/2}$的情形作一比较.

6. 考虑用

$$A = \begin{bmatrix} 10 & 3 & -2 & 1 \\ 3 & 10 & 0 & 9 \\ -2 & 0 & 10 & 4 \\ 1 & 9 & 4 & 10 \end{bmatrix}$$

给出的矩阵 $A \in M_4$. 证明 A 是半正定矩阵但 $|A|$ 不是半正定矩阵.

7. 设 $K(x, y)$是有限区间$[a, b]$上的连续积分核. 证明 $K(x, y)$是半正定核当且仅当对点集$\{x_i\}_{i=1}^n \subset [a, b]$的所有选择和所有 $n=1, 2, \cdots$，矩阵$[K(x_i, x_j)] \in M_n$ 是半正定矩阵. **提示：**为了证明矩阵条件是充分的，考虑对积分的 Riemann 和逼近

$$\int_a^b\int_a^b K(x,y)f(x)\bar{f}(y)\mathrm{d}x\mathrm{d}y \cong \sum_{i,j=1}^n K(x_i,x_j)f(x_i)\bar{f}(x_j)\Delta x_i\Delta x_j.$$

为了证明矩阵条件是必要的，考虑函数

462

$$f(x)\equiv\sum_{i=1}^n a_i\delta_\varepsilon(x-x_i)$$

其中 $\delta_\varepsilon(x)$是"近似 δ 函数"，它是连续的，和非负的，在区间$[-\varepsilon, \varepsilon]$外恒为零，且适合

$$\int_{-\infty}^{\infty}\delta_\varepsilon(x)\mathrm{d}x=1,$$

然后让 $\varepsilon\to 0$.

8. 试用习题 7 和 Schur 乘积定理证明半正定积分核的(逐点)乘积是半正定的. 这个证明方法是比较基本的，它不需要积分方程理论中的 Mercer 定理.

9. 证明，函数 $\phi\in C(\mathbf{R})$是正定函数[(7.1)节习题 8]当且仅当 $K(x, y)\equiv\phi(x-y)$是半正定积分核.

10. 证明，两个正定函数 $\phi_1(x)$，$\phi_2(x)$的积$(\phi_1\phi_2)(x)$是正定函数.

11. 说明为什么所有函数

(a) $\dfrac{\sin(Tx)}{Tx}=\dfrac{1}{2T}\displaystyle\int_{-T}^{T}e^{itx}\mathrm{d}t,\quad T>0,$

(b) $e^{-x^2}=\dfrac{1}{2\sqrt{\pi}}\displaystyle\int_{-\infty}^{\infty}e^{-t^2/2}e^{itx}\mathrm{d}t,$

(c) $e^{-|x|}=\dfrac{1}{\pi}\displaystyle\int_{-\infty}^{\infty}\dfrac{e^{itx}}{1+t^2}\mathrm{d}t,$

以及它们彼此间的所有乘积都是正定函数.

12. 利用 11(c)给出(7.2)节习题 12 中的矩阵是正定矩阵的另一个证明.

13. 如果 $A=[a_{ij}]\in M_n$ 是半正定矩阵，证明矩阵$[a_{ij}/(i+j)]$也是半正定矩阵. **提示**：(7.1)节习题 17.

14. 设 $A\in M_n$ 是半正定矩阵. 证明，$x\in\mathbf{C}^n$ 适合 $x^*Ax=0$ 当且仅当 $Ax=0$. 如果 A 仅仅是 Hermite 矩阵，用例子说明可能有 $x^*Ax=0$ 而 $Ax\neq 0$. **提示**：将 A 写成 $A=U\Lambda U^*$，于是 $x^*Ax=0$ 当且仅当 $\Sigma\lambda_i|z_i|^2=0$，其中 $z=U^*x$.

15.(顶点在原点 0 的)凸锥是这样一个凸集 S，使得射线$\{\lambda x:\lambda\geqslant 0\}\subset S$对所有 $x\in S$ 都成立. 凸锥 S 的一条射线$\{\lambda x:\lambda\geqslant 0\}$是一条极射线，是指若对于 $0<\alpha<1$ 和 y，$z\in S$ 有 $x=\alpha y+(1-\alpha)z$ 仅当 y 和 z 都在该射线上，等价地，凸锥的一条射线是一条极射线，是指把它从该锥中删掉后所得到的锥还是凸的. 证明，M_n 中由半正定矩阵组成的凸锥中的射线$\{\lambda A:\lambda\geqslant 0\}$是一条极射线当且仅当 A 的秩为 1. 于是定理(7.5.2)是说，每一个半正定矩阵是位于极射线上的诸矩阵的一个凸组合. **提示**：(a)如果 $x\in\mathbf{C}^n$ 是非零的，且 $xx^*=\alpha A+(1-\alpha)B$ 对某个 $\alpha\in(0, 1)$以及半正定矩阵 A，$B\in M_n$ 成立，设$\{x_2, \cdots, x_n\}\in\mathbf{C}^n$ 是一个适合 $x^*x_k=0$(其中 $k=2, \cdots, n$)的标准正交组. 于是 $0=x_k^*Ax_k=x_k^*Bx_k$，因而根据习题 14 每个 x_k 在 A 和 B 的零空间中. 由此得出，A 和 B 的秩都是 1，并且各自都是 xx^* 的一个正纯量倍数. (b)如果 $A\in M_n$ 是半正定的，且其秩 $k\geqslant 2$，利用定理(7.5.2)记 $A=B+C$，其中 $B=vv^*$，$v\neq 0$，rank $C\geqslant$

463

1，且 $Cv=0$. 由此得出，C 不是 B 的一个纯量倍数，因而 A 不在极射线上.

7.6 相合：乘积和同时对角化

与正实数的乘法不同，普通的矩阵乘法不总保持正定性. 两个 Hermite 矩阵的乘积甚至可以不是 Hermite 矩阵(只有当它们可交换时，乘积才是 Hermite 矩阵)，且乘积诱导的二次型可以不是非负的. 我们特别把这一节的重点放在正定矩阵上；关于 Hermite 矩阵的更一般的结果见(4.5)节.

7.6.1 例 设 $A=\begin{bmatrix}5&-2\\-2&1\end{bmatrix}$和 $B=\begin{bmatrix}2&1\\1&1\end{bmatrix}$，$A$ 和 B 是正矩定阵. 但 $AB=\begin{bmatrix}8&3\\-3&-1\end{bmatrix}$不是对称矩阵. $H(AB)=\begin{bmatrix}8&0\\0&-1\end{bmatrix}$甚至连正定矩阵都不是.

但是，正定矩阵的普通乘积至少还保留了一点正性. 我们的讨论要说明某些涉及矩阵的和与积的有用技巧.

7.6.2 定义 我们知道，两个矩阵 A，B *相合，是指存在非奇异矩阵 $C\in M_n$ 使得 $B=C^*AC$.

注意，和相似一样，*相合是等价关系. 有时在复的情形采用术语共轭相合，以便把它和实相合区别开来.

464

7.6.3 定理 正定矩阵 $A\in M_n$ 与 Hermite 矩阵 $B\in M_n$ 的乘积是可对角化矩阵，它的所有特征值都是实数. 矩阵 AB 与 B 有相同数目的正特征值，负特征值和零特征值. 此外，任意只具有实特征值的可对角化矩阵是一个正定矩阵与一个 Hermite 矩阵的乘积.

证明： 对于前一部分，注意到 $A^{-1/2}ABA^{1/2}=A^{1/2}BA^{1/2}$，于是后一个矩阵相似于 AB，因而与它恰好有相同的特征值. 因为 $A^{1/2}$是 Hermite 矩阵，所以矩阵 $A^{1/2}BA^{1/2}$相合于 B. 因此，根据 Sylvester 惯性定理(4.5.8)，B 的特征值与 $A^{1/2}BA^{1/2}$的，因而与 AB 的特征值有相同的符号集. 此外，因为 $A^{1/2}BA^{1/2}$是 Hermite 矩阵，它可对角化，因而 AB 也一定可对角化. 关于后一个论断，假定 $C\in M_n$ 是可对角化矩阵，且只有实特征值：$C=SDS^{-1}$，其中 D 是实对角矩阵. 则 $C=S(S^*S^{*-1})DS^{-1}=(SS^*)(S^{-1*}DS^{-1})=AB$，其中 $A\equiv SS^*$ 是正定矩阵而 $B=S^{-1*}DS^{-1}$是 Hermite 矩阵. □

两个矩阵可经相似同时对角化是不常有的，需要较强的附加条件：交换性. 但是，两个 Hermite 矩阵可经共同的*相合同时对角化所需要的假设条件就很弱. 经*相合同时对角化对应于用诸变量的一个线性变换把两个 Hermite 二次型变换成诸平方项的线性组合. 下面的结果是经典的；关于综合的结果见(4.5.15).

7.6.4 定理 设 A，$B\in M_n$ 是两个 Hermite 矩阵，且假定存在 A 和 B 的一个实线性组合可以是正定矩阵，则存在非奇异矩阵 $C\in M_n$ 使得 C^*AC 和 C^*BC 都是对角矩阵.

证明： 假定对某个 α，$\beta\in\mathbf{R}$，$P=\alpha A+\beta B$ 是正定矩阵. α 和 β 中至少有一个必定非零，所以可以假定 $\beta\neq 0$. 但是，因为 $B=\beta^{-1}(P-\alpha A)$，如果能证明 A 和 P 可经*相合同时对对角化，则可以得出，A 和 B 也可经*相合同时对角化. 由(7.2.7)可知，P *相合于单位矩阵，即存在

某个非奇异矩阵 $C_1\in M_n$ 使得 $C_1^*PC_1=I$. 因为 $C_1^*AC_1$ 是 Hermite 矩阵，所以存在酉矩阵 U 使得 $U^*C_1^*AC_1U=D$ 是对角矩阵. 令 $C\equiv C_1U$，则有 $C^*PC=I$ 和 $C^*AC=D$，并且 $C^*BC=\beta^{-1}(I-\alpha D)$是对角矩阵. □

[465]

这个结果最常见的应用是针对力学中的经典情形的，其中，两个实对称二次型是给定的且有一个是正定的.

7.6.5 推论 如果 $A\in M_n$ 是正定矩阵，且 $B\in M$ 是 Hermite 矩阵，则存在非奇异矩阵 $C\in M_n$，使得 C^*BC 是对角矩阵且 $C^*AC=I$.

练习 试求诸变量的一个变换使得两个二次型 $5x^2-2xy+y^2$ 和 $x^2+2xy-y^2$ 都是平方项的加权和.

对于一个是正定矩阵，另一个是(复)对称矩阵的矩阵偶，也有类似的结果. 这个结果也可以归并到(4.5.15)中.

7.6.6 定理 如果 $A\in M_n$ 是正定矩阵，且 $B\in M_n$ 是复对称矩阵，则存在一个非奇异矩阵 C 使得C^*AC 和C^TBC 都是对角矩阵.

证明：我们选取非奇异矩阵 $C_1\in M_n$ 使得 $C_1^*AC_1=I$. 于是 $C_1^TBC_1$ 是对称矩阵，因而根据 Takagi 分解(4.4.4)，存在酉矩阵 U 使得 $U^T(C_1^TBC_1)U=D$，其中 D 是对角矩阵. 于是也有 $U^*C_1^*AC_1U=I$，这样，可以取 $C\equiv C_1U$. □

这个结果可应用于复变函数论；关于单叶函数的 Grunsky 不等式是由正定 Hermite 矩阵和复对称矩阵所诱导的二次型之间的不等式.

下面的结果是(7.6.5)的直接应用.

7.6.7 定理 函数 $f(A)=\log\det A$ 是 M_n 中正定 Hermite 矩阵组成的凸集上的严格凹函数.

证明：对任意两个给定的正定矩阵 A，$B\in M_n$，必须证明，

$$f(\alpha A+(1-\alpha)B)\geqslant \alpha f(A)+(1-\alpha)f(B) \tag{7.6.8}$$

对所有 $\alpha\in(0,1)$成立，而其中等式成立当且仅当 $A=B$. 利用(7.6.5)写出 $A=CIC^*$ 和 $B=C\Lambda C^*$，其中，$C\in M_n$ 是某个非奇异矩阵，$\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，且所有 $\lambda_i>0$. 于是，

[466]

$$\begin{aligned}f(\alpha A+(1-\alpha)B)&=f(C[\alpha I+(1-\alpha)\Lambda]C^*)=f(CC^*)+f(\alpha I+(1-\alpha)\Lambda)\\&=f(A)+f(\alpha I+(1-\alpha)\Lambda),\end{aligned}$$

且

$$\begin{aligned}\alpha f(A)+(1-\alpha)f(B)&=\alpha f(A)+(1-\alpha)f(C\Lambda C^*)\\&=\alpha f(A)+(1-\alpha)[f(CC^*)+f(\Lambda)]\\&=\alpha f(A)+(1-\alpha)f(A)+(1-\alpha)f(\Lambda)\\&=f(A)+(1-\alpha)f(\Lambda).\end{aligned}$$

因此只需证明，对所有 $\alpha\in(0,1)$，$f(\alpha I+(1-\alpha)\Lambda)\geqslant(1-\alpha)f(\Lambda)$对具有正对角元的任意对角矩阵 Λ 成立. 但这容易从对数函数本身的严格凹性推出，因为

$$f(\alpha I+(1-\alpha)\Lambda)=\log\prod_{i=1}^{n}[\alpha+(1-\alpha)\lambda_i]=\sum_{i=1}^{n}\log[\alpha+(1-\alpha)\lambda_i]$$

$$\geqslant \sum_{i=1}^{n}[\alpha\log 1+(1-\alpha)\log\lambda_i]$$

$$=(1-\alpha)\sum_{i=1}^{n}\log\lambda_i=(1-\alpha)\log\prod_{i=1}^{n}\lambda_i$$

$$=(1-\alpha)\log\det A=(1-\alpha)f(A).$$

这个不等式中的等式成立，当且仅当每个 $\lambda_i=1$，而这能成立当且仅当 $A=I$ 且 $B=CIC^*=A$. □

定理(7.6.7)常常采用下述形式，它是对不等式(7.6.8)取幂得到的，它对正定矩阵的凸组合是正定矩阵，因而一定是非奇异矩阵的事实给出了数量表示.

7.6.9 推论 设 A，$B\in M_n$ 是正定矩阵，且设 $0<\alpha<1$. 则

$$\det[\alpha A+(1-\alpha)B]\geqslant[\det A]^\alpha[\det B]^{1-\alpha},$$

其中等式成立当且仅当 $A=B$.

习题

1. 假定 $A\in M_n$ 适合 $A^*=S^{-1}AS$，其中 $S\in M_n$ 是正定矩阵. 证明 A 可对角化且 A 的所有特征值都是实数. **提示**：考虑 $AS=SA^*$. 证明 AS 是 Hermite 矩阵，然后利用(7.6.3). [467]

2. 证明 $f(A)=\operatorname{tr}A^{-1}$ 是关于正定矩阵的严格凸函数. **提示**：(7.6.7)的证明.

3. 如果 $A\in M_n$ 是半正定矩阵，如何推广(7.6.3)? 证明，AB 的特征值还是实数，且 AB 的正特征值和负特征值不会比 B 的多，不过它可能有更多的零特征值.

4. 如果 $B\in M_n$ 不是 Hermite 矩阵，(7.6.3)可以推广到什么程度?

5. 试用例子说明，可能两个 Hermite 矩阵可经相合同时对角化，但它们不满足(7.6.4)的假设条件.

6. 设 A，$B\in M_2$ 是给定的 Hermite 矩阵. 就 A 和 B 的特征值而论，AB 的两个特征值的实部的所有可能符号是什么? 你能把它推广到 M_n 吗?

7. 设 A，$B\in M_n$ 是 Hermite 矩阵，且 A 是正定矩阵. 试用(7.6.5)证明，$A+B$ 是正定矩阵，当且仅当 $A^{-1}B$ 的每个特征值大于-1. **提示**：$A+B=A(I+A^{-1}B)$.

8. 设 $H\in M_n$ 是 Hermite 矩阵，将 H 写成 $H=A+iB$，其中 A，$B\in M_n(\mathbf{R})$. 验证 A 是对称矩阵而 B 是斜对称矩阵，因而 B 的特征值是纯虚的，且成共轭对出现. 试证 H 是正定矩阵当且仅当 A 是正定矩阵且 $iA^{-1}B$ 的每个特征值大于-1. **提示**：利用 $x^*Hx=x^*Ax$ 对所有 $x\in\mathbf{R}^n$ 成立的事实. 利用习题 7. 如果 A 是正定矩阵，证明，如果 λ 是 $iA^{-1}B$ 的特征值，则$-\lambda$ 亦是它的特征值. 由此得出，H 是正定矩阵，当且仅当 A 是正定矩阵且 $iA^{-1}B$ 的每个特征值位于区间$(-1，1)$内，并且 $iA^{-1}B$ 的特征值成对$\{-\lambda，\lambda\}$出现. 由此得出 $0\leqslant\det iA^{-1}B<1$，因而 $\det B<\det A$，这是 H. P. Robertson 的不等式. 现在记 $H=A+iB=A(I+iA^{-1}B)$，然后证明，如果 H 是正定矩阵，则 $\det H=\det A\det(I+iA^{-1}B)$且 $0<\det(I+iA^{-1}B)<1$. 由此得出，如果 H 是正定矩阵，则 $\det H\leqslant\det A$，这是 O. Taussky 的不等式.

9. 由定理(4.1.7)可知，矩阵 $A\in M_n$ 是两个 Hermite 矩阵的乘积，当且仅当 A 相似于实矩阵. 试用(7.6.3)证明，$A\in M_n$ 是两个正定的 Hermite 矩阵的乘积，当且仅当 A 可对角化且

[468] 只有正特征值. **提示**：关于逆命题，考虑 $A=SAS^{-1}=SS^*(S^{-1})^*AS^{-1}$.

10. 如果 $A, B\in M_n$ 是正定矩阵，则我们知道，乘积 AB 是正定矩阵当且仅当 AB 是 Hermite 矩阵. 证明相同的结论对三个正定矩阵的乘积也成立；也就是说，如果 $A, B, C\in M_n$ 是正定矩阵，则乘积 $S=ABC$ 是正定矩阵，当且仅当它是 Hermite 矩阵. **提示**：记 $S=(AB)C=EC$，根据习题 9，其中 E 有 n 个正特征值. 利用(7.6.3)证明，如果 S 是 Hermite 矩阵，则 $E=SC^{-1}$与 S 有相同数目的正特征值.

11. 对习题 10 中的结果的下述另一个证明作详细的论述：设 $S(\alpha)\equiv[(1-\alpha)C+\alpha A]BC$，其中 $0\leqslant\alpha\leqslant 1$. 如果 $S(1)$是 Hermite 矩阵，又因为 $S(0)=CBC$ 自然是 Hermite 矩阵，所以所有 $S(\alpha)$都是 Hermite 矩阵. 证明所有 $S(\alpha)$都是非奇异的，因为$(1-\alpha)C+\alpha A$ 是非奇异的. $S(\alpha)$的诸特征值连续地依赖 α，当 $\alpha=0$ 时所有特征值是正的，因为所有 $S(\alpha)$是非奇异的，故它无零特征值. 由此得出 $S(1)$的所有特征值是正的.

进一步阅读 关于取自各种正定类的矩阵所作乘积的其他结果，以及关于多个正定矩阵之积的较早结果，其有关的资料可参看 C. S. Ballantine and C. R. Johnson"Accretive Matrix Products,"*Lin. Multilin. Alg.* 3(1975)，169-185.

7.7 半正定次序关系

因为 Hermite 矩阵是实数的推广，而正定矩阵是正实数的推广，自然要问，在 Hermite 矩阵类中是否有不等式关系或(偏)序关系的适当概念.

7.7.1 定义 设 $A, B\in M_n$ 是 Hermite 矩阵. 如果 $A-B$ 是半正定矩阵，我们就用 $A\geqslant B$ 表示，类似地 $A>B$ 表示 $A-B$ 是正定矩阵.

练习 证明上述的不等概念与矩阵的相等概念是一致的，即证明 $A\geqslant B$ 和 $B\geqslant A$ 蕴涵 $A=B$.

练习 证明，关系$\geqslant$是传递的和自反的，但它不是全序；即存在 Hermite 矩阵 $A, B\in M_n$ [469] 使得 $A\geqslant B$ 和 $B\geqslant A$ 都不成立. 这样的关系称为偏序.

实线性空间上的偏序常常定义如下：确定某个特殊的闭凸锥，并且说一个元素大于或等于另一个元素，是指它们的差位于这个特殊的锥中. 在这种情形，$n\times n$ Hermite 矩阵的集合是实线性空间，而半正定矩阵的集合是闭凸锥. 这显然是 **R** 自身为实线性空间而非负实数集为闭凸锥这一熟知情形的推广. 不过在 **R** 上给出的是"普通"的(全)序(而不仅仅是偏序).

矩阵间的各种其他的"不等"概念(其中最值得注意的是按照分量确定实矩阵的大小)可以用类似的方法来定义：把矩阵的一个锥看作非负实数的推广，并且说 A"大于或等于"B，是指它们的差 $A-B$ 位于这个锥中. 一般，这样一些不同的"不等"概念可以通过上下文来区别，不过它们的效用取决于与实数集的类似推广到什么程度以及这个"不等"概念与其他不等关系(例如特征值、行列式等之间的不等关系)有多么密切的关系.

注意，A 是半正定矩阵，当且仅当 $A\geqslant 0$；而 A 是正定矩阵，当且仅当 $A>0$，其中 0 是与 A 同阶的零矩阵.

练习 用例子说明半正定偏序关系在下述情形与实数的全序关系不同：如果 $A\geqslant B$ 且 A 不

等于 B，则得不出 $A\succ B$.

下面，说明半正定次序关系的某些性质，其中每一个可以看作实数的普通次序关系的推广. 其类似程度一般是很强的.

7.7.2 论断 如果 A、$B\in M_n$ 是 Hermite 矩阵，则

$$A\succeq B \text{ 蕴涵 } T^*AT\succeq T^*BT$$

对所有 $T\in M_{n,m}$ 成立；如果 $m\leqslant n$，且 $T\in M_{n,m}$ 有秩 m，则还有

$$A\succ B \text{ 蕴涵 } T^*AT\succ T^*BT.$$

证明： 如果 $A-B$ 是半正定矩阵，则 $y^*(A-B)y\geqslant 0$ 对所有 $y\in \mathbf{C}^n$ 成立. 于是 $x^*(T^*AT-T^*BT)x=(Tx)^*(A-B)(Tx)\geqslant 0$ 对所有 $x\in\mathbf{C}^m$ 成立，而这表示 T^*AT-T^*BT 是半正定矩阵. 因此 $T^*AT\succeq T^*BT$. 注意，这推广了(7.1.6)，而其证明本质上是相同的. □ 470

练习 验证上述第二个论断以完成证明.

7.7.3 定理 设 A，$B\in M_n$ 是 Hermite 矩阵，又假定 A 是正定矩阵且 B 是半正定矩阵. 则 $A\succeq B$ 当且仅当 $\rho(BA^{-1})\leqslant 1$，而 $A\succ B$ 当且仅当 $\rho(BA^{-1})<1$.

证明： 根据(7.6.5)，可以求出非奇异矩阵 $C\in M_n$ 使得 $A=CIC^*$ 和 $B=CDC^*$，其中 $D=\mathrm{diag}(d_1, d_2, \cdots, d_n)$ 是对角矩阵. 于是 $A\succeq B$ 当且仅当 $C[I-D]C^*\succeq 0$，这又当且仅当 $d_i\leqslant 1$ 对所有 $i=1, 2, \cdots$ 成立. 但是因为 $BA^{-1}=CDC^*C^{*-1}C^{-1}=CDC^{-1}$，所以 BA^{-1} 的特征值正好是 $d_1, d_2, \cdots, d_n$[根据(7.6.3)它们都是非负的]，且所有 $d_i\leqslant 1$ 当且仅当 $\rho(BA^{-1})\leqslant 1$. 仔细分析刚用过的那些不等式便可得出后一个论断. □

7.7.4 推论 如果 A，$B\in M_n$ 是正定矩阵，则

(a) $A\succeq B$ 当且仅当 $B^{-1}\succeq A^{-1}$；

(b) 如果 $A\succeq B$，则 $\det A\geqslant \det B$ 且 $\mathrm{tr}\, A\geqslant \mathrm{tr}\, B$；

(c) 更一般地，如果 A 和 B 的各相应特征值按同一(递增或递减)顺序排列，则 $\lambda_k(A)\geqslant \lambda_k(B)$ 对所有 $k=1, 2, \cdots, n$ 成立.

证明： 我们知道 $A\succeq B$ 当且仅当 $\rho(BA^{-1})\leqslant 1$. 但是 $\rho(BA^{-1})=\rho(A^{-1}B)$，而(7.7.3)说明 $\rho(A^{-1}B)\leqslant 1$ 当且仅当 $B^{-1}\succeq A^{-1}$. 如果 $A\succeq B$，则 $\rho(BA^{-1})\leqslant 1$，又由(7.6.3)可知，BA^{-1} 的所有特征值是非负的，所以得知它们必定位于区间(0，1]内. 另一方面，它们的乘积至多为 1，所以 $\det(BA^{-1})\leqslant 1$，因而 $\det A\geqslant \det B$. 在(7.7.3)的证明中，已知 $A=CC^*$ 和 $B=CDC^*$，其中 $C=[c_{ij}]\in M_n$，$D=\mathrm{diag}(d_1, d_2, \cdots, d_n)\in M_n$，且对所有 $i=1, 2, \cdots, n$，$0\leqslant d_i\leqslant 1$. 不难算出，

$$\mathrm{tr}\, A=\mathrm{tr}\, CC^*=\sum_{i,j=1}^{n}|c_{ij}|^2,$$

且

$$\mathrm{tr}\, B=\mathrm{tr}\, CDC^*=\mathrm{tr}\, DC^*C=\sum_{i,j=1}^{n}d_i|c_{ij}|^2$$

$$\leqslant \sum_{i,j=1}^{n} |c_{ij}|^2 = \operatorname{tr} A.$$

最后一个论断(它蕴涵行列式不等式和迹不等式，对此，我们已经给出了两个无关的证明)可直接从 Hermite 矩阵的有序特征值的 Courant-Fischer 变分特征推出，并且包括在推论(4.3.3)中. □

471

练习 如果 $A>B>0$，证明 $\det A>\det B$ 和 $\operatorname{tr} A>\operatorname{tr} B$.

当把关于分块矩阵的逆的形式(0.7.3)限制到 Hermite 矩阵的情形时，便得出下述有用的公式：

$$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}^{-1} = \begin{bmatrix} (A-BC^{-1}B^*)^{-1} & A^{-1}B(B^*A^{-1}B-C)^{-1} \\ (B^*A^{-1}B-C)^{-1}B^*A^{-1} & (C-B^*A^{-1}B)^{-1} \end{bmatrix}. \tag{7.7.5}$$

在这个公式中，假定 A 和 C 是方阵且需求逆的矩阵是非奇异的.

如果矩阵$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$是正定矩阵，则$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}^{-1}$存在且为正定矩阵. 于是从(7.7.5)和(7.1.2)推出$(A-BC^{-1}B^*)^{-1}$和$A-BC^{-1}B^*$是正定矩阵. 类似地，$C-B^*A^{-1}B$，A 和 C 是正定矩阵. 因此，如果分块 Hermite 矩阵$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$是正定的，则有

$$A>0,\quad C>0,\quad A>BC^{-1}B^* \text{ 和 } C>B^*A^{-1}B.$$

7.7.6 定理 假定一个 Hermite 矩阵块分成

$$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix},$$

其中 A 和 C 是方阵. 这个矩阵是正定的当且仅当 A 是正定矩阵且$C>B^*A^{-1}B$. 此外，这个条件等价于一定有 $\rho(B^*A^{-1}BC^{-1})<1$.

证明：这两个条件的必要性已在上面做了论述. 关于充分性，假定 A 是正定矩阵且 $C>B^*A^{-1}B$，对 $X=-A^{-1}B$，算出

$$\begin{bmatrix} I & 0 \\ X^* & I \end{bmatrix}\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}\begin{bmatrix} I & X \\ 0 & I \end{bmatrix}=\begin{bmatrix} A & 0 \\ 0 & C-B^*A^{-1}B \end{bmatrix},$$

因为右边是正定矩阵，所以

$$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$$

的正定性可从所给出的相合及(7.1.6)或(7.7.2)推出. 把(7.7.3)应用于不等式 $C>B^*A^{-1}B$ 便得到后一个论断. □

472

练习 如果$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}>0$，证明 $\det C>\det B^*A^{-1}B$ 和 $\det A>\det BC^{-1}B^*$. 当 $B\in M_{n,1}$ 时，这说明什么？证明，如果 B 是方阵，则 $\det A\det C\geqslant |\det B|^2$.

练习 假定 $A\in M_n$，$C\in M_m$ 和 $B\in M_{n,m}$，又假定 A 和 C 都是正定矩阵. 证明$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}\geqslant 0$

当且仅当 $\rho(B^*A^{-1}BC^{-1})\leqslant 1$.

(7.7.6)中的分块正定矩阵与出现在复变函数和调和分析中的某些双线性不等式有关，这些不等式都具有正定偏序的某些性质.

7.7.7 定理 设 $A\in M_n$ 和 $C\in M_m$ 是正定矩阵，且设 $B\in M_{n,m}$. 则下列条件等价：

(a) $(x^*Ax)(y^*Cy)\geqslant |x^*By|^2$ 对所有 $x\in \mathbf{C}^n$ 和所有 $y\in \mathbf{C}^m$ 成立；

(b) $x^*Ax+y^*Cy\geqslant 2|x^*By|$ 对所有 $x\in \mathbf{C}^n$ 和所有 $y\in \mathbf{C}^m$ 成立；

(c) $\rho(B^*A^{-1}BC^{-1})\leqslant 1$；

(d) $\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}\geqslant 0$.

证明：我们来证明，(a)蕴涵(b)，(b)蕴涵(c)，(c)蕴涵(a)；已经知道(c)和(d)是等价的. 如果(a)成立，则由算术—几何平均值不等式有

$$\frac{1}{2}(x^*Ax+y^*Cy)\geqslant (x^*Ax)^{1/2}(y^*Cy)^{1/2}\geqslant |x^*By|,$$

所以(b)成立. 如果假定(b)成立，则

$x^*Ax+y^*Cy=(A^{1/2}x)^*(A^{1/2}x)+(C^{1/2}y)^*(C^{1/2}y)\geqslant 2|x^*By|$，因而对每个 $x\in \mathbf{C}^n$ 和每个 $y\in \mathbf{C}^m$ 有

$$x^*x+y^*y\geqslant 2|(A^{-1/2}x)^*B(C^{-1/2}y)|=2|x^*A^{-1/2}BC^{-1/2}y|.$$

如果在这个不等式中令 $x\equiv A^{-1/2}BC^{-1/2}y$，便得

$$y^*C^{-1/2}B^*A^{-1}BC^{-1/2}y+y^*y\geqslant 2|y^*C^{-1/2}B^*A^{-1}BC^{-1/2}y|.$$

因为矩阵 $C^{-1/2}B^*A^{-1/2}BC^{-1/2}$ 是半正定的，这等价于对所有 $y\in \mathbf{C}^m$ 有

$$y^*y\geqslant y^*C^{-1/2}B^*A^{-1}BC^{-1/2}y.$$

如果选取 y 为 $C^{-1/2}B^*A^{-1}BC^{-1/2}$ 的特征向量，这个不等式说明，(一定非负)的相应特征值不大于1，于是得知谱半径至多是1；即 $1\geqslant\rho(C^{-1/2}B^*A^{-1}BC^{-1/2})=\rho(B^*A^{-1}BC^{-1})$，因而(c)成立. 最后，如果(c)成立，则对任意 $x\in \mathbf{C}^n$ 和任意 $y\in \mathbf{C}^m$ 有

473

$$\begin{aligned}|x^*(A^{-1/2}BC^{-1/2}y)|^2 &\leqslant \|x\|_2^2\,\|A^{-1/2}BC^{-1/2}y\|_2^2\\ &=(x^*x)(y^*C^{-1/2}B^*A^{-1}BC^{-1/2}y)\leqslant (x^*x)(y^*y),\end{aligned}$$

其中 $\|x\|_2\equiv(x^*x)^{1/2}$ 是 Euclid 范数. 如果我们现在作代换 $x\to A^{1/2}x$ 和 $y\to C^{1/2}y$，则对所有 $x\in \mathbf{C}^n$ 和所有 $y\in \mathbf{C}^m$，有

$$|x^*By|^2\leqslant (x^*Ax)(y^*Cy).$$

□

对于由(7.7.5)产生的另一种不等式，我们考虑能应用于正定矩阵的两种可能运算：基于指定的指标集选定主子矩阵与求矩阵的逆. 我们知道这两种运算保持正定性，但是按两种可能的顺序实施这两运算的结果之间存在什么关系呢？研究结果表明，这两种运算“除了相差一个不等关系以外是可交换的”.

7.7.8 定理 假定 $P\in M_n$ 是正定矩阵，且设 $S\subset\{1, 2, \cdots, n\}$ 是一个指标集，则

$$P^{-1}(S)\geqslant [P(S)]^{-1},$$

其中，这个不等式左边是划去 P^{-1} 的标号为 S 的诸行和诸列后所确定的 P^{-1} 的主子矩阵，而右边是 P 的相应主子矩阵的逆.

证明：因为正定矩阵的集合在置换相合下封闭，我们可以假定

$$P=\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$$

及 $P(S)=A$. 于是 $P^{-1}(S)=(A-BC^{-1}B^*)^{-1}$ 且 $[P(S)]^{-1}=A^{-1}$. 由于 $C>0$（这是因为 $P>0$），所以有 $BC^{-1}B^*\geqslant 0$ 和

$$A\geqslant A-BC^{-1}B^*\geqslant 0.$$

于是所断言的不等式可由(7.7.4a)推出. □

定理(7.7.8)可以解释为"一个正定矩阵的主子矩阵的逆小于或等于该矩阵的逆的相应主子矩阵."

(7.7.8)的一个应用是从矩阵的 Kronecker 乘积中特殊选择一个主子矩阵来产生其 Hadamard 乘积. 如果 A，$B\in M_n$，又如果 $S=\{1, n+2, 2n+3, 3n+4, \cdots, n^2\}$，则 $A\circ B=(A\otimes B)(S)$. 若 A 和 B 可逆，则 $A\otimes B$ 可逆且 $(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$. 因此，如果 A 和 B 是正定矩阵且把(7.7.8)用于 $P=A\otimes B$，则 $A^{-1}\circ B^{-1}=(A^{-1}\otimes B^{-1})(S)=(A\otimes B)^{-1}(S)\geqslant[(A\otimes B)(S)]^{-1}=(A\circ B)^{-1}$. 如果取 $B=A$，这说明 $A^{-1}\circ A^{-1}\geqslant(A\circ A)^{-1}$. 但是，如果取 $B=A^{-1}$，这便说明，当 A 是正定矩阵时，$A^{-1}\circ A\geqslant(A\circ A^{-1})^{-1}=(A^{-1}\circ A)^{-1}$.

471

这后一个不等式说明，$A^{-1}\circ A$ 优于它自己的逆. 关于 $A^{-1}\circ A$，这指的是什么？如果 C 是正定矩阵，且 $C=U\Lambda U^*$，$\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，其中所有 $\lambda_i>0$，则 $C\geqslant C^{-1}$ 当且仅当 $\lambda_i\geqslant 1$，因而 $C\geqslant I\geqslant C^{-1}$. 我们把这些结论总结为

7.7.9 定理　设 A，$B\in M_n$ 是正定矩阵，则

(a) $A^{-1}\circ B^{-1}\geqslant(A\circ B)^{-1}$；

(b) $A^{-1}\circ A^{-1}\geqslant(A\circ A)^{-1}$；

(c) $A^{-1}\circ A\geqslant I\geqslant(A^{-1}\circ A)^{-1}$.

因为 $A^{-1}A=I$，所以(c)的前一部分说明 $A^{-1}\circ A\geqslant A^{-1}A$；即在这种情形，Hdamard 乘法优于普通乘法.

习题

1. 一般，设 A，$B\in M_n$ 是 Hermite 矩阵，且 $A\geqslant B$，证明，如果 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$ 是 A 的有序特征值，且 $\mu_1\leqslant\mu_2\leqslant\cdots\leqslant\mu_n$ 是 B 的有序特征值，则 $\lambda_i\geqslant\mu_i$，$i=1, 2, \cdots, n$. 但是用例子说明逆命题不总是成立的.

2. 如果 A_1，A_2，B_1，$B_2\in M_n$ 都是 Hermite 矩阵，证明，如果 $A_1\geqslant B_1$，且 $A_2\geqslant B_2$，则 $A_1+A_2\geqslant B_1+B_2$.

3. 设 A，B，$C\in M_n$ 是 Hermite 矩阵；假定 $A\geqslant B$ 和 $C\geqslant 0$. 证明 $A\circ C\geqslant B\circ C$.

4. 设 A，B，C，$D\in M_n$ 是 Hermite 矩阵，且假定 $A\geqslant B\geqslant 0$ 和 $C\geqslant D\geqslant 0$. 利用前一个习题证明 $A\circ C\geqslant B\circ D\geqslant 0$.

5. 如果 A，$B\in M_n$ 是使得 $A\geqslant B$ 的 Hermite 矩阵，又如果 $J\subset\{1, 2, \cdots, n\}$ 是任意指标

集．证明 $A(J) \geqslant B(J)$.

6. 说明(7.7.6)推广了 $n=2$ 时的(7.2.5).

7. 如果 $C \in M_1$，(7.7.6)说的是什么？如何为一个正定矩阵增添一行和一列且仍保持正定性？

475

8. 证明，(7.7.8)的不等式是严格的，当且仅当 $P(S, S')$ 有满行秩，而当 $P(S, S')=0$ 时恰好等式成立．这里，$P(S, S')$ 是从 P 中划去标号为 S 的诸行和标号为 S' 的诸列后得到的 P 的子矩阵．**提示**：证明 $\operatorname{rank}[P^{-1}(S)-P(S)^{-1}]=\operatorname{rank} P(S, S')$.

9. 如果 $A \in M_n$ 是 Hermite 矩阵，证明 $I \geqslant A$ 当且仅当 A 的所有特征值小于或等于 1.

10. 试用(7.7.7)给出(7.4)节习题 11 的另一个解法．**提示**：证明，若 $B>0$，则 $\begin{bmatrix} B & I \\ I & B^{-1} \end{bmatrix} \geqslant 0$.

11. 考虑 $A=C$ 时的(7.7.7)．证明下列条件等价：

(a) $(x^* A x)(y^* A y) \geqslant |x^* B y|^2$ 对所有 $x, y \in \mathbf{C}^n$ 成立；

(b) $x^* A x + y^* A y \geqslant \frac{1}{2}|x^* B y|^2$ 对所有 $x, x \in \mathbf{C}^n$ 成立；

(c) $\rho(B^* A^{-1} B A^{-1}) \leqslant 1$；

(d) $x^* A x \geqslant |x^* B x|$ 对所有 $x \in \mathbf{C}^n$ 成立.

12. 证明，如果 $A \in M_n$ 是可逆对称矩阵，则 $A^{-1} \circ A$ 的所有行和等于 1．**提示**：考察 A^{-1} 各元的代数余子式．这样，如果 A 是实正定矩阵，证明，即使 $A^{-1} \circ A \geqslant I$，也不会有 $A^{-1} \circ A > I$.

13. 如果 $A^{(k)}$ 表示 A 的 Hadamard k 次幂，又如果 $A \in M_n$ 是正定矩阵，证明 $(A^{-1})^{(k)} \geqslant (A^{(k)})^{-1}$ 对所有 $k=1, 2, \cdots$ 成立.

进一步阅读　关于(7.7.7)的背景材料以及其他资料可参看 C. FitzGerald and R. Horn, "On the Structure of Hermitian-Symmetric Inequalities," *J. London Math. Soc.* 15(2)(1997), 419-430. 与(7.7.8)，(7.7.9)有关的其他资料也可参看 C. Johnson, "Partitioned and Hadamard Product Matrix Inequalities," *J. Research NBS* 83(1978), 585-591.

7.8 关于正定矩阵的不等式

下面，要讨论这样一些不等式，它们所涉及的量与一个或多个正定矩阵相关联．这些不等式与上一节所介绍的不等式是有区别的，尽管关于前者的一些例子与关于后者的一些例子常常是有联系的．例如，$A \geqslant B \geqslant 0$ 蕴涵 $\det A \geqslant \det B$．正定矩阵有一些涉及行列式，特征值和其他量的不等式．在这一节，我们考察某些不等式，它们不一定来自矩阵不等式.

476

关于正定矩阵的基本行列式不等式是 Hadamard 不等式．许多其他的不等式用不同的方式推广了这一结果.

7.8.1 定理(Hadamard 不等式)　如果 $A=[a_{ij}] \in M_n$ 是半正定矩阵，则

$$\det A \leqslant \prod_{i=1}^{n} a_{ii}.$$

此外，如果 A 是正定矩阵，则等式成立当且仅当 A 是对角矩阵.

证明：如果 A 是奇异矩阵，就没有什么可证的，因此假定 A 是非奇异矩阵且它的所有 $a_{ii}\neq 0$. 定义 $d_i\equiv a_{ii}^{-1/2}$，且设 $D=\operatorname{diag}(d_1, d_2, \cdots, d_n)$. 因为 $\det DAD\leqslant 1$ 当且仅当 $\det A\leqslant a_{11}a_{22}\cdots a_{nn}$，所以只要假定 A 的每个对角元等于 1 就可以了. 如果 $\lambda_1, \cdots, \lambda_n$ 是 A 的(必定为正的)特征值，则有

$$\det A=\prod_{i=1}^{n}\lambda_i\leqslant\left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i\right)^n=\left(\frac{1}{n}\operatorname{tr}A\right)^n=1$$

这个不等式可从关于非负实数的算术一几何平均值不等式推出. 算术一几何平均值不等式中等式成立当且仅当所有 $\lambda_i=1$，但是，因为 A 是 Hermite 矩阵，因而可对角化，所以，这种情况能出现，当且仅当 $A=I$. 因此，当 A 是正定矩阵时，在原不等式中等式成立当且仅当 A 是对角矩阵. □

关于一般方阵的另一个行列式不等式等价于(7.8.1)，它也叫做 Hadamard 不等式. 几何上 $|\det A|$ 是 n 维平行六面体的体积，它的各生成边是由 A 的各行(或各列)给出的. 当各生成边相互正交时，这个体积最大，而这时其体积是各边长的乘积. Hadamard 不等式是这个几何不等式的代数描述.

7.8.2 推论(Hadamard 不等式) 对于任意矩阵 $B=[b_{ij}]\in M_n$，有

$$|\det B|\leqslant\prod_{i=1}^{n}\left(\sum_{j=1}^{n}|b_{ij}|^2\right)^{1/2}$$

和

$$|\det B|\leqslant\prod_{j=1}^{n}\left(\sum_{i=1}^{n}|b_{ij}|^2\right)^{1/2}.$$

477 此外，等式成立当且仅当 B 的各行(相应地，各列)相互正交.

证明：如果 B 是奇异矩阵，那就没有什么要证的. 如果 B 是非奇异矩阵，把(7.8.1)应用正定矩阵 $A=BB^*$，然后取方根. 第一个不等式右边是 A 的各对角元的乘积的方根，而左边是 $\det A$ 的方根. 当 A 是对角矩阵时，即在(7.8.1)中等式成立的情形恰好 B 的各行互相正交. 把第一个不等式应用于 B^* 便可推出第二个不等式. □

练习 我们已经从(7.8.1)推导出(7.8.2). 现在证明(7.8.1)可由(7.8.2)推出. **提示**：如果 A 是正定矩阵，则存在唯一的正定矩阵 B 使 $B^2=A$. 把(7.8.2)应用于 B 和它的平方.

练习 你能利用 Hadamard 不等式(及其变形)给出

$$\det\begin{bmatrix}1 & 1 & 1\\ 1 & -1 & -1\\ 1 & -1 & 1\end{bmatrix}$$

的最大界吗？

对关于正定矩阵的 Hadamard 不等式作出改进的两个推广应归功于 Fischer 和 Szasz. 在 Fischer 不等式中，互余主子矩阵所起的作用相当于对角元在 Hadamard 不等式中所起的作用.

7.8.3 定理(Fischer 不等式) 假定

$$P=\begin{bmatrix}A & B\\ B^* & C\end{bmatrix}$$

是正定矩阵，其中子块 A 和 C 是非空方阵(参看本节 14 题)．则

$$\det P \leqslant (\det A)(\det C)$$

证明：设 $X = -A^{-1}B$，然后算出

$$\det P = \det\begin{bmatrix} I & 0 \\ X^* & I \end{bmatrix}\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}\begin{bmatrix} I & X \\ 0 & I \end{bmatrix} = \det\begin{bmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{bmatrix}$$

$$= (\det A)(\det[C - B^*A^{-1}B]) \leqslant (\det A)(\det C)$$

后一不等式用到了(7.7.6)和(7.7.4b)以确保 $\det C \geqslant \det(C - B^*A^{-1}B)$，这是因为，$C \geqslant C - B^*A^{-1}B \geqslant 0$． □ [478]

练习 试从 Fischer 不等式推导出 Hadamard 不等式(7.8.1)．同时，试对于比(7.8.3)中的分块(两个主子矩阵)要细但又不如(7.8.1)中分块(n 个主子矩阵)那么细的 P 的分块，提出并叙述 Fischer 不等式．注意，在这种情形，Fischer 不等式的右边小于或等于 Hadamard 不等式的右边．因此，关于加细的各种分块的 Fischer 不等式给出关于 $\det P$ 的上界的一个单调不减序列．

存在另一种不等式，它给出关于行列式的一系列上界，且包括 Hadamard 上界．设 $P_k(A)$ 表示 A 的所有 $k \times k$ 主子式$\left[\text{共有}\binom{n}{k}\text{个}\right]$的乘积．我们注意到，$P_n(A) = \det A$，而 $P_1(A) = a_{11}a_{22}\cdots a_{nn}$．

7.8.4 定理(Szasz 不等式) 如果 $A \in M_n$ 是正定矩阵，则对所有 $k = 1, 2, \cdots, n-1$ 有

$$P_{k+1}(A)^{\binom{n-1}{k}^{-1}} \leqslant P_k(A)^{\binom{n-1}{k-1}^{-1}}$$

证明：因为 A^{-1} 的各对角元正好是 A 的各 $(n-1)\times(n-1)$ 主子式与 $\det A$ 之比，把(7.8.1)直接应用于正定矩阵 A^{-1} 便推出

$$\frac{1}{\det A} = \det A^{-1} \leqslant \frac{P_{n-1}(A)}{(\det A)^n},$$

因而

$$P_n(A)^{n-1} = (\det A)^{n-1} \leqslant P_{n-1}(A).$$

对这个不等式的两边取 $(n-1)$ 次方根便得出 Szasz 不等式组中 $k = n-1$ 的情形．余下情形可以归纳地导出．例如，对于 $k = n-2$ 的情形，我们把每个 $(n-1)\times(n-1)$ 主子矩阵看作一个起始矩阵，然后应用上面的不等式得到

$$P_{n-1}(A)^{n-2} \leqslant P_{n-2}(A)^2$$

因为 A 的每个 $(n-2)\times(n-2)$ 主子矩阵作为 A 的某个 $(n-1)\times(n-1)$ 主子矩阵的主子矩阵出现两次．对两边取 $(n-1)(n-2)$ 次方根得 $k = n-2$ 的情形，而余下的情形可用同样的方式推出． □

练习 证明 Szasz 不等式蕴涵 Hadamard 不等式(7.8.1)．其中等式的情形是什么？ [479]

7.8.5 论断 设 $A \in M_n$ 是半正定矩阵，且定义

$$\alpha(A) \equiv \begin{cases} \dfrac{\det A}{\det A_{11}}, \text{如果 } A_{11} \text{ 是正定矩阵；} \\ 0, \text{否则；} \end{cases}$$

其中 A_{11} 是划去 A 的第 1 行和第 1 列而得的 A 的 $(n-1)\times(n-1)$ 主子矩阵. 设 $E_{11}\in M_n$ 是其 1.1 元为 1，而所有其余元为 0 的矩阵. 则对所有 $t\leqslant\alpha(A)$，$A-tE_{11}$ 是半正定矩阵，而对任意 $t>\alpha(A)$，$A-tE_{11}$ 不是半正定矩阵；特别是，$A-\alpha(A)E_{11}$ 是半正定矩阵.

证明：只要考虑 A 是正定阵的情形即可. 将(7.2.5)应用于各"尾随"主子式. 注意到 $A-tE_{11}$ 的前 $n-1$ 个尾随主子式与 A 的相同，且 $\det(A-tE_{11})=\det A-t\det A_{11}$. □

练习 给出(7.8.5)的证明细节.

练习 试用(7.8.5)归纳证明 Hadamard 不等式(7.8.1).

Hadamard 不等式(7.8.1)也可以用 Hadamard 乘积叙述为

$$(\det A)\prod_{i=1}^{n}1\leqslant\det A\circ I.$$

下述定理是 Oppenheim 的不等式(被 Schur 强化了)，它说明在上述不等式中单位矩阵的作用没有什么特殊的，从而推广了 Hadamard 不等式.

7.8.6 定理(Oppenheim 不等式) 如果 A，$B\in M_n$ 是半正定矩阵，则

$$(\det A)\prod_{i=1}^{n}b_{ii}\leqslant\det A\circ B.$$

证明：我们对 n 采用归纳法. 对 $n=1$，结论是明显的. 如果 $n\geqslant2$，且对阶数至多为 $n-1$ 的所有矩阵，结论成立，则由归纳假设可知

480

$$(\det A_{11})\prod_{i=2}^{n}b_{ii}\leqslant\det A_{11}\circ B_{11},$$

其中的记号与(7.8.5)中相同，注意 $A_{11}\circ B_{11}=(A\circ B)_{11}$. 因为 $A-\alpha E_{11}$ 是半正定矩阵，由此得知 $(A-\alpha E_{11})\circ B$ 是半正定矩阵，因而

$$0\leqslant\det(A-\alpha E_{11})\circ B=(\det A\circ B)-\alpha b_{11}(\det A_{11}\circ B_{11}).$$

由此可推出

$$\det A\circ B\geqslant\alpha b_{11}\det A_{11}\circ B_{11}\geqslant\alpha b_{11}(\det A_{11})\prod_{i=2}^{n}b_{ii}=(\det A)\prod_{i=1}^{n}b_{ii}.$$ □

练习 如果 A，$B\in M_n$ 是正定矩阵，证明，

$$(\det A)(\det B)\leqslant\det A\circ B$$

进而证明

$$(\det A)(\det B)\leqslant(\det A)\prod_{i=1}^{n}b_{ii}\leqslant\det A\circ B\leqslant\prod_{i=1}^{n}a_{ii}\prod_{i=1}^{n}b_{ii}.$$

练习 如果 $A\in M_n$ 是正定矩阵，证明 $\det A\circ A^{-1}\geqslant1$.

一种类型不同的行列式不等式适应于其 $H(A)$ 是正定矩阵的非 Hermite 矩阵 A. 它可以看作关于复数的不等式 $|z|\geqslant|\mathrm{Re}\,z|$ 的推广.

7.8.7 定理(Ostrowski-Taussky) 如果 $A\in M_n$ 使 $H(A)\equiv(A+A^*)/2$ 为正定矩阵，则

$$\det H(A)\leqslant|\det A|.$$

等式成立当且仅当 A 是 Hermite 矩阵.

证明：设 $S(A)=(A-A^*)/2$，因而 $A=H(A)+S(A)$. 于是所要证的不等式可述为

$$|\det[I+H(A)^{-1}S(A)]|\geqslant 1.$$

但是 $H(A)^{-1}S(A)$ 相似于斜 Hermite 矩阵

$$H(A)^{-1/2}S(A)H(A)^{-1/2},$$

因而它只有纯虚特征值. 因此只要注意到 $|1+it|\geqslant 1$ 对任意实数 t 成立就可以了. 如果 it_1，it_2，…，it_n 是 $H(A)^{-1}S(A)$ 的特征值，则

481

$$|\det[I+H(A)^{-1}S(A)]|=\prod_{j=1}^{n}|1+it_j|\geqslant 1.$$

另外，等式成立当且仅当所有 $t_j=0$，它等价于 $S(A)=0$，这是因为斜 Hermite 矩阵是可对角化的. □

含有两个正定矩阵之和的重要行列式不等式属于 Minkowski. 它的证明与上述结果类似.

7.8.8 定理(Minkowski 不等式) 如果 A，$B\in M_n$ 是正定矩阵，则

$$[\det(A+B)]^{1/n}\geqslant(\det A)^{1/n}+(\det B)^{1/n}.$$

证明：注意到所要证的不等式的两边是同次齐次的，用$(\det A^{-1/2})^{1/n}$乘其左边和右边. 于是，不失一般性，可以假定 $A=I$，且必须证明

$$[\det(I+B)]^{1/n}\geqslant 1+(\det B)^{1/n}.$$

如果 $0<\lambda_1\leqslant\cdots\leqslant\lambda_n$ 是 B 的特征值，则所要证的不等式等价于

$$\prod_{i=1}^{n}(1+\lambda_i)\geqslant(1+\sqrt[n]{\lambda_1\cdots\lambda_n})^n,$$

只要直接计算不等式两边，然后用算术-几何平均值不等式逐项进行比较便可证明这个不等式. □

练习 给出(7.8.8)的证明细节，并且证明等式成立当且仅当 $B=cA$ 对某个 $c\geqslant 0$ 成立.

习题

1. 不等式(7.8.2)说明，行列式的大小可以用其各行的 l_2 范数的乘积来估计. 试把它同如下结果[见(6.1)节习题 3]作一比较：行列式的大小可以用其各行的 l_1 范数的乘积来估计. 几何上，这每一个界表示什么？还有其他这样的界吗？试一试 l_∞.

482

2. (7.8.2)的左边在 B 的左酉乘法下不变，而(7.8.1)的左边在 A 的酉相似下不变，但是，两者的右边都无相应的不变性. 什么时候两个右边达到极小？什么时候它们达到极大？能用这种方式得到较好的界吗？

3. 试用 Fischer 不等式验证 Hadamard 不等式(7.8.2)的下述分块形式的推广：设 $A=[A_{ij}]$是 $nk\times nk$ 分块复矩阵，使得每个子块 $A_{ij}\in M_k$. 于是

$$|\det A|\leqslant\left[\prod_{i=1}^{n}\left(\sum_{j=1}^{n}\|A_{ij}\|_2^2\right)^{1/2}\right]^k.$$

除了这里所采用的谱范数以外，还可以用其他矩阵范数吗？

4. 试确定(7.8.6)中相等的情形.

5. 设 A，$B\in M_n$ 是正定矩阵，试证

$$\det A\circ B+(\det A)(\det B)\geqslant(\det A)\prod_{i=1}^{n}b_{ii}+(\det B)\sum_{i=1}^{n}a_{ii}.$$

说明这强化了(7.8.6). **提示**：证明

$$\alpha(A\circ B)\geqslant\alpha(A)b_{11}+\alpha(B)a_{11}-\alpha(A)\alpha(B),$$

然后把它应用于普通的归纳假设.

6. 证明，可以进一步扩展上一个习题中的不等式得到下述不等式

$$\det A\circ B+(\det A)(\det B)\frac{\det A_{11}\circ B_{11}}{(\det A_{11})(\det B_{11})}$$

$$\geqslant(\det A)\prod_{i=1}^{n}b_{ii}+(\det B)\prod_{i=1}^{n}a_{ii}+(\det A)b_{11}(\det B_{11})\left(\frac{a_{22}\cdots a_{nn}}{\det A_{11}}-1\right)$$

$$+\left[\det(B)a_{11}(\det A_{11})\left(\frac{b_{22}\cdots b_{nn}}{\det B_{11}}-1\right)\right].$$

7. 如果 $A\in M_n$ 有正定 Hermite 部分 $H=(A+A^*)/2$，且 $n>1$，证明(7.8.7)中的不等式可以强化为

$$\det H(A)+|\det S(A)|\leqslant|\det A|.$$

什么情形下可取等式？**提示**：你必须证明

$$|\det[I+H^{-1}(A)S(A)]|\geqslant 1+|\det H(A)^{-1}S(A)|,$$

483 这等价于有

$$\prod_{j=1}^{n}|1+it_j|\geqslant 1+\prod_{j=1}^{n}|t_j|$$

证明

$$\prod_{j=1}^{n}|1+it_j|^2=1+\sum_{j=1}^{n}t_j^2+\cdots+\prod_{j=1}^{n}t_j^2$$

$$\geqslant 1+n\prod_{j=1}^{n}t_j+\left(\prod_{j=1}^{n}t_j\right)^2\geqslant\left(1+\prod_{j=1}^{n}|t_j|\right)^2.$$

你可以进一步强化这个不等式吗？注意：所论证的不等式关于复数的一个自然形式应该是 $|z|\geqslant|\operatorname{Re}z|+|\operatorname{Im}z|$，而所论证的不等式可以看作它的推广；证明这个自然不等式不成立(因而假定 $n>1$)，由此可见要证的行列式不等式有点出乎意料之外.

8. 如果 $A,B\in M_n$ 是正定矩阵，证明 $\det(A+B)\geqslant\det A+\det B$.

9. 试用 Minkowski 不等式证明 Fischer 不等式. **提示**：把 Minkowski 不等式应用于两个正定矩阵

$$\begin{bmatrix}A & B\\ B^* & C\end{bmatrix}\quad 和\quad\begin{bmatrix}I & 0\\ 0 & -I\end{bmatrix}\begin{bmatrix}A & B\\ B^* & C\end{bmatrix}\begin{bmatrix}I & 0\\ 0 & -I\end{bmatrix}=\begin{bmatrix}A & -B\\ -B^* & C\end{bmatrix}.$$

10. 一个正定矩阵 $P\in M_n$ 可以分解成 $P=LL^*$，其中 L 是只有正对角元的下三角矩阵(7.2.9). 试用这一事实证明 Fischer 不等式.

11. 设 $A,B\in M_n$ 是半正定矩阵. 如果 A 和 B 是非奇异的，证明 $A\circ B$ 是非奇异的(且是正定的). 如果 $A\circ B$ 是奇异的，证明 A、B 至少有一个是奇异的. 这与(7.5)节的不等式 rank

$A \circ B \leqslant (\operatorname{rank} A)(\operatorname{rank} B)$有何关系?

12. 证明，如果$A=[a_{ij}]\in M_3$是一个具有实元素的矩阵，且$|a_{ij}|\leqslant 1$，则$|\det A|\leqslant 3\sqrt{3}$. 同时证明这个界是不能达到的. **提示:**

$$\frac{\partial}{\partial a_{ij}}(\det A)=(-1)^{i+j}A_{ij} \quad 且 \quad \frac{\partial^2}{\partial a_{ij}^2}(\det A)\equiv 0,$$

其中A_{ij}是A划去A的第i行和第j列的行列式. 如果$A_{ij}=0$，则$\det A$与a_{ij}的值无关，因而a_{ij}可取± 1. 如果$A_{ij}\neq 0$，则当$0<a_{ij}<1$时，$\det A$关于a_{ij}没有极值. 因此，在所有$a_{ij}=\pm 1$的给定的约束内，$|\det A|$达到它的最大值. 对于$n=3$，只有有限个这样的矩阵. 一般地，对于$n>3$，其结果又如何? 如果A有复元素，则对解析函数应用最大值原理(最大模定理)证明，在集合$\{A\in M_n$: 所有$|a_{ij}|\leqslant 1\}$内部$|\det A|$不可能有最大值. [484]

13. 如果$A=[a_{ij}]\in M_n$且$K=\max\{|a_{ij}|\}$，试用 Hadamard 不等式证明，$|\det A|\geqslant K^n n^{n/2}$.

14. 设$A\in M_n$是正定矩阵，又设$\alpha\subseteq N=\{1,\cdots,n\}$是一个指标集. Fischer 不等式可以叙述为$\det A\leqslant \det A(\alpha)\det A(\alpha')$，其中$\alpha'$是$\alpha$关于$N$的补集. 这个结果的一个推广常称为 Hadamard-Fischer 不等式，那就是

$$\det A(\alpha\cup\beta)\leqslant \frac{\det A(\alpha)\det A(\beta)}{\det A(\alpha\cap\beta)}, \tag{7.8.9}$$

它对正定 Hermite 矩阵A和所有指标集α，$\beta\subseteq N$成立. 通常规定$\det A(\phi)\equiv 1$. 试仅用 Fischer 不等式和(0.8.4)中的第二个公式证明 Hadamard-Fischer 不等式. **提示:** 不失一般性，假定$\alpha\cup\beta=N$，且把 Fisher 不等式应用于$A^{-1}(\alpha'\cup\beta')$. 然后把(0.8.4)应用于每个子式.

15. 利用正定 Hermite 矩阵可以写成LL^*且L是非奇异下三角矩阵的事实(7.2.9)给出 Hadamard-Fischer 不等式(7.8.9)的直接证明. **提示:** 假定$1\leqslant j<k<n$，不失一般性，假定$\alpha=\{1,\cdots,k\}$和$\beta=\{1,\cdots,j,k+1,\cdots,n\}$. 然后考虑$A$和$L$的一个对应的$3\times 3$子块.

16. 设$A\in M_n$是正定矩阵. 试用 Hadamard Fischer 不等式(7.8.9)证明

$$\det A\leqslant \frac{\sum_{i=1}^{n-1}\det A(\{i,i+1\})}{\prod_{i=2}^{n-1}a_{ii}}.$$

17. 设$A\in M_n$是正定矩阵. 证明

$$\det A=\min\left\{\sum_{i=1}^{n}v_i^* A v_i : \{v_1,\cdots,v_n\}\subset \mathbf{C}^n \text{ 是标准正交组}\right\}.$$

提示: 设$V=[v_1\cdots v_n]\in M_n$且把(7.8.1)应用于$\tilde{A}\equiv V^* AV$.

18. 设$A\in M_n$是正定矩阵，且设$\{u_1,\cdots,u_n\}\subset\mathbf{C}^n$是标准正交组. 试用习题 17 证明，$\{u_1,\cdots,u_n\}$是$A$的特征向量且$\{u_1^* Au_1,\cdots,u_n^* Au_n\}$是$A$的相应特征值，当且仅当 [485]

$$\det A=\prod_{i=1}^{n}u_i^* Au_i.$$

19. 如果$A\in M_n$是正定矩阵，证明

$$n(\det A)^{1/n} = \min\{\operatorname{tr} AB : B \in M_n \text{ 是正定矩阵且 } \det B = 1\}.$$

提示：记 $A=U\Lambda U^*$，其中，$\Lambda=\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$，所有 $\lambda_i>0$，且 $U\in M_n$ 是酉矩阵，于是 $\operatorname{tr} AB=\operatorname{tr}\Lambda(U^* BU)$. 然后利用算术-几何平均值不等式和 Hadamard 不等式(7.8.1)证明

$$\frac{1}{n}\sum_{i=1}^{n}\lambda_i b_{ii} \geqslant \left(\prod_{i=1}^{n}\lambda_i b_{ii}\right)^{1/n} = \left(\det A\prod_{i=1}^{n} b_{ii}\right)^{1/n} \geqslant [\det A]^{1/n},$$

且等式可以成立.

20. 试用习题 19 中的拟线性化证明 Minkowski 不等式(7.8.8)

21. 设 $A\in M_n$ 是半正定矩阵，且把 A 写成分块形式如下：

$$A = \begin{bmatrix} a_{11} & x^* \\ x & \tilde{A} \end{bmatrix}$$

试用(4.1)节习题 15 中关于行列式的简化公式和(7.2)节习题 11 证明

$$\det A = a_{11}\det \tilde{A} - x^*(\operatorname{adj}\tilde{A})x \leqslant a_{11}\det\tilde{A}.$$

试用归纳法和这个不等式给出 Hadamard 不等式(7.8.1)的另一个证明.

进一步阅读 关于定理(7.8.7)的其他信息可参看 A. M. Ostrowski and O. Taussky, "On the Variation of the Determinant of a Positive Definite Matrix," *Proc. Kon. Nederl. Acad. Wetensch. Amsterdam*, Ser. A, 54 (1951), 383-385. 有一类不等式(当 A 是正定矩阵时)把 $\det A$ 与 A 的其他广义矩阵函数(见 0.3.2)联系起来，同时也推广了 Hadamard 不等式(7.8.1)，其有关的内容可参看 I. Schur, "Über endliche Gruppen und Hermitesche Formen," *Math. Z.* 1 (1918), 184-207.

486

第8章　非负矩阵

8.0　导引

假定有 $n\geqslant 2$ 个城市 C_1，…，C_n，它们之间的人口流动情况如下所述：对所有 $i\neq j$，每天上午 8：00 同时从 j 城到 i 城去的流动人口的正常比率是 a_{ij}，j 城的流动人口仍留在 j 城的比率是 a_{jj}．因此，如果用 $p_i^{(m)}$ 表示第 m 天 i 城的人口数，则有第 m 天与第 $m+1$ 天之间的人口分布递归关系

$$p_i^{(m+1)} = a_{i1}p_1^{(m)} + \cdots + a_{in}p_n^{(m)}, \quad i = 1,\cdots,n, \quad m = 0,1,\cdots.$$

如果用 $A=[a_{ij}]$ 表示由人口流动系数组成的 $n\times n$ 矩阵，且用 $p^{(m)}=[p_i^{(m)}]$ 表示人口分布向量，则

$$p^{(m+1)} = Ap^{(m)} = AAp^{(m-1)} = \cdots = A^{m+1}p^{(0)}, \quad m = 0,1,\cdots,$$

其中 $p^{(0)}$ 是初始人口分布向量．因为系数 a_{ij} 表示流动人口比率，所以，对每个 $j=1$，1，…，n，有 $0\leqslant a_{ij}\leqslant 1$，且 $\sum_{i=1}^{n} a_{ij} = 1$.

为了对城市服务以及资金投入作出切合实际的长远规划，行政人员想知道，从现在到遥远的将来，总的人口 $p = \sum_{i=1}^{n} p_i^{(0)}$ 将如何分布；即他们想知道，对大的 m，$p^{(m)}$ 的渐近变化过程．但是，因为 $p^{(m)}=A^m p^{(0)}$，所以显然必须考察 $A^{(m)}$ 的渐近变化过程．

例如，详细考虑 $n=2$ 的情形．因为 $a_{11}+a_{21}=1=a_{12}+a_{22}$，如果记 $a_{21}=\alpha$ 和 $a_{12}=\beta$，则有

$$A = \begin{bmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{bmatrix},$$

我们想知道，对大的 m，A^m 等于什么？如果 A 是可对角化的，则可以直接计算 A^m．从计算 A 的特征值开始：$\lambda_2=1$ 和 $\lambda_1=1-\alpha-\beta$．因为 $0\leqslant\alpha$，$\beta\leqslant 1$，所以 $\lambda_2=1\geqslant|\lambda_1|=|1-\alpha-\beta|$，于是有 $1=|\lambda_2|=\rho(A)$ 且 A 的谱半径是 A 的一个特征值．另外，除了在 $\alpha=\beta=0$ 的平凡情形（在这种情形，A 是可约矩阵）以外，我们看到 $\lambda_2=\rho(A)$ 是 A 的单重特征值．

如果 $\alpha+\beta\neq 0$，相应的特征向量是 $x=[\beta,\ \alpha]^T$（对于 $\lambda_2=1$）和 $z=[1,\ -1]^T$（对于 λ_1），所以在这种情形 A 是可对角化的且 $A=S\Lambda S^{-1}$，其中

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{bmatrix}, \quad S = \begin{bmatrix} \beta & 1 \\ \alpha & -1 \end{bmatrix} \quad \text{和} \quad S^{-1} = \frac{1}{\alpha+\beta}\begin{bmatrix} 1 & 1 \\ \alpha & -\beta \end{bmatrix}.$$

注意，如果 A 是不可约的，则特征向量 x 的分量是非负的和正的．

如果 α 和 β 不都是 1，则 $|\lambda_1|=|1-\alpha-\beta|<1$，因而当 $m\to\infty$ 时 $\lambda_1^m\to 0$．因此，在这种情形，有

$$\lim_{m\to\infty} A^m = S(\lim_{m\to\infty}\Lambda^m)S^{-1} = S\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}S^{-1} = \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \beta \\ \alpha & \alpha \end{bmatrix},$$

因而平衡人口分布是

$$\lim_{m\to\infty} p^{(m)} = \frac{1}{\alpha+\beta}\begin{bmatrix}\beta & \beta\\ \alpha & \alpha\end{bmatrix}\begin{bmatrix}p_1^{(0)}\\ p_2^{(0)}\end{bmatrix} = \frac{1}{\alpha+\beta}\begin{bmatrix}\beta\\ \alpha\end{bmatrix}.$$

注意，平衡分布完全与初始分布无关．矩阵 A^m 趋近于一个极限，这个极限的各列与相应于特征值 1(它是 A 的谱半径)的特征向量 x 成正比，而极限人口分布与这同一个特征向量成正比．

对于上面没有处理的两种例外情形不难逐一分析．如果 $\alpha=\beta=0$，则 $A=I$，$\lim\limits_{m\to\infty} A^m=I$，且 $\lim\limits_{m\to\infty} p^{(m)}=p^{(0)}$，所以极限分布与初始分布不是无关的．

如果 $\alpha=\beta=1$，则 $A=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$且两个城市逐日调换它们的全部人口．$A$ 的幂不趋近于一个极限，并且，如果初始人口分布是不平衡的，则人口分布也无极限．但是，在"均值平衡"的意义下，它们都可达到一个极限，即

488

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=1}^{m}A^k=\begin{bmatrix}.5 & .5\\ .5 & .5\end{bmatrix}\quad 和 \quad \lim_{m\to\infty}\frac{1}{m}\sum_{k=1}^{m}p^{(k)}=\frac{p_1^{(0)}+p_2^{(0)}}{2}\begin{bmatrix}1\\ 1\end{bmatrix}.$$

总之，从这个例子得知，$\rho(A)=1$，并且：

1. 谱半径 $\rho(A)=1$ 本身是 A 的特征值，而它不只是特征值的绝对值．

2. 相应于特征值 $\rho(A)$ 的特征向量 x 可以取非负分量，如果 A 是不可约的，则分量是正的．

3. 如果 A 的所有元都是正的，则 $\rho(A)$ 是具有严格最大模的单重特征值．

4. 如果 A 的所有元是正的，则 $\lim\limits_{m\to\infty}[A/\rho(A)]^m$ 存在，且是秩 1 矩阵，它的所有列与特征向量 x 成正比．

5. 在所有情形 $\lim\limits_{m\to\infty}(1/m)\sum\limits_{k=1}^{m}(A/\rho(A))^k$ 存在．

实际上这些结论一般对 $n\geqslant 2$ 是成立的，不过用上面所述的简单而直接的方法不可能分析研究一般情形．例如，当 $n\geqslant 2$ 时，即使 A 的所有元素都是正的，A 也不一定可对角化．这就需要新的工具，这正是本章的其余部分要阐述的内容．

习题

1. 证明，矩阵 $A=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$有谱半径 1，但是当 $m\to\infty$时 A^m 无界．

2. 考虑矩阵

$$A_\varepsilon=\begin{bmatrix}\dfrac{1}{1+\varepsilon} & \dfrac{1}{1+\varepsilon}\\ \dfrac{\varepsilon^2}{1+\varepsilon} & \dfrac{1}{1+\varepsilon}\end{bmatrix},\quad \varepsilon>0.$$

(a) 证明，$\lambda_2=1$ 是 A_ε 的单重特征值，$\rho(A)=\lambda_2=1$，且 $1>|\lambda_1|$．(b)证明

$$x=\frac{1}{1+\varepsilon}\begin{bmatrix}1\\ \varepsilon\end{bmatrix}\quad 和 \quad y=\frac{1+\varepsilon}{2\varepsilon}\begin{bmatrix}\varepsilon\\ 1\end{bmatrix}$$

分别是 A_ε 和 A_ε^T 相应于特征值 $\lambda=1$ 的特征向量．(c)直接计算 A_ε^m，$m=1,2,\cdots$．(d)证明

489

$$\lim_{m\to\infty}A_\varepsilon^m=\frac{1}{2}\begin{bmatrix}1 & \varepsilon^{-1}\\ \varepsilon & 1\end{bmatrix}.$$

(e)计算 xy^T，并作出说明. (f)如果 $\varepsilon\to 0$，会出现什么情形？**提示**：令 $B_\varepsilon=(1+\varepsilon)A_\varepsilon$，然后如正文中那样对角化 B_ε.

3. 从人口自由流动的观点来说明，由城市间的人口流动系数组成的一般矩阵是不可约的意味着什么.

4. 对于在本节中所讨论过的两个城市的例子，试证明，在 $\alpha=\beta=0$ 以及 $\alpha+\beta\neq 0$ 的情形下 $\lim_{m\to\infty}(1/m)\sum_{k=1}^{m}A^k$ 存在. 在每种情形下的极限各是什么？

进一步阅读 关于正定矩阵和非负矩阵的性质的丰富信息，以及许多理论与应用的参考文献可见[BPl]和[Sen]. 在[Var]中，有一个特别注重数值分析应用的关于非负矩阵的诸多结果的综述.

8.1 非负矩阵——不等式及其推广

设 $B=[b_{ij}]\in M_{n,r}$ 和 $A=[a_{ij}]\in M_{n,r}$. 记

$B\geqslant 0$，如果所有 $b_{ij}\geqslant 0$；

$B>0$，如果所有 $b_{ij}>0$；

$A\geqslant B$，如果 $A-B\geqslant 0$；

$A>B$，如果 $A-B>0$.

类似地可定义反向关系 $\leqslant$ 和 $<$. 我们定义 $|A|=[|a_{ij}|]$. 如果 $A\geqslant 0$，则称 A 是非负矩阵，又如果 $A>0$，则称 A 是正矩阵. 下列简单事实可直接由定义推出.

练习 设 $A, B\in M_{n,r}$. 证明：

(8.1.1) $|A|\geqslant 0$ 对每个 A 成立；$|A|=0$ 当且仅当 $A=0$.

(8.1.2) $|aA|=|a||A|$ 对所有 $a\in\mathbf{C}$ 成立.

(8.1.3) $|A+B|\leqslant|A|+|B|$.

(8.1.4) 如果 $A\geqslant 0$ 且 $A\neq 0$，则当 n 或 r 大于 1 时 $A>0$ 不一定成立.

(8.1.5) 如果 $A\geqslant 0$，$B\geqslant 0$，且 a，$b\geqslant 0$，则 $aA+bB\geqslant 0$.

(8.1.6) 如果 $A\geqslant B$ 且 $C\geqslant D$，则 $A+C\geqslant B+D$.

(8.1.7) 如果 $A\geqslant B$ 且 $B\geqslant C$，则 $A\geqslant C$.

490

练习 现在假定 A，B，C，$D\in M_n$，且 x，$y\in\mathbf{C}^n$. 证明：

(8.1.8) $|Ax|\leqslant|A||x|$.

(8.1.9) $|AB|\leqslant|A||B|$.

(8.1.10) $|A^m|\leqslant|A|^m$ 对所有 $m=1, 2, \cdots$ 成立.

(8.1.11) 如果 $0\leqslant A\leqslant B$ 且 $0\leqslant C\leqslant D$，则 $0\leqslant AC\leqslant BD$.

(8.1.12) 如果 $0\leqslant A\leqslant B$，则 $0\leqslant A^m\leqslant B^m$ 对所有 $m=1, 2, \cdots$ 成立.

(8.1.13) 如果 $A\geqslant 0$，则 $A^m\geqslant 0$；如果 $A>0$，则 $A^m>0$ 对所有 $m=1, 2, \cdots$ 成立.

(8.1.14) 如果 $A>0$，$x\geqslant 0$，且 $x\neq 0$，则 $Ax>0$.

(8.1.15) 如果 $A\geqslant 0$，$x>0$ 且 $Ax=0$，则 $A=0$.

(8.1.16) 如果 $|A|\leqslant|B|$，则 $\|A\|_2\leqslant\|B\|_2$.

(8.1.17) $\|A\|_2=\||A|\|_2$.

显然，最后两个论断对任意绝对范数都成立，Frobenius 范数(l_2 范数)只是其中一个例子.这些简单关系首先可应用于关于谱半径的不等式.

8.1.18 定理 设 $A, B\in M_n$. 如果 $|A|\leqslant B$，则 $\rho(A)\leqslant\rho(|A|)\leqslant\rho(B)$.

证明：对每个 $m=1, 2, \cdots$，根据(8.1.10)和(8.1.12)，有 $|A^m|\leqslant|A|^m\leqslant B^m$. 因此，根据(8.1.16)和(8.1.17)，对所有 $m=1, 2, \cdots$，不等式

$\|A^m\|_2\leqslant\||A|^m\|_2\leqslant\|B^m\|_2$ 和 $\|A^m\|_2^{1/m}\leqslant\||A|^m\|_2^{1/m}\leqslant\|B^m\|_2^{1/m}$ 成立. 如果现在设 $m\to\infty$ 且运用(5.6.14)，便导出 $\rho(A)\leqslant\rho(|A|)\leqslant\rho(B)$. □

8.1.19 推论 设 $A, B\in M_n$. 如果 $0\leqslant A\leqslant B$，则 $\rho(A)\leqslant\rho(B)$.

8.1.20 推论 设 $A\in M_n$. 如果 $A\geqslant 0$，又如果 $\tilde{A}$ 是 A 的任一主子矩阵，则 $\rho(\tilde{A})\leqslant\rho(A)$. 特别是，$\max_{i=1,\cdots,n}a_{ii}\leqslant\rho(A)$.

证明：设 $1\leqslant r\leqslant n$，且设 $\tilde{A}$ 是 A 的 $r\times r$ 主子方阵. 设 $\hat{A}$ 表示把 $\tilde{A}$ 的各元排放在它们原来的位置(作为 A 的元)而把 0 元放在其余位置所构成的 $n\times n$ 矩阵. 于是 $\rho(\tilde{A})=\rho(\hat{A})$ 且 $0\leqslant\hat{A}\leqslant A$，所以根据推论(8.1.9)可知，$\rho(\tilde{A})=\rho(\hat{A})\leqslant P(A)$. □

对于一个不一定是 Hermite 矩阵的谱半径来说，上述推论中的下界 $a_{ii}\leqslant\rho(A)$ 是得到的第一个非平凡下界，不过 A 为非负的假定是本质的.

491

练习 试构造一个相似于 $\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$ 的矩阵，且不含零元. 它的谱半径是什么？它是非负矩阵吗？关于推论(8.1.20)的后一部分，这说明什么问题？

练习 证明，如果 $A, B\in M_n$，且 $0\leqslant A<B$，则 $\rho(A)<\rho(B)$. **提示**：存在某个 $\alpha>1$ 使得 $0\leqslant A\leqslant\alpha A<B$. 如果 $\rho(A)\neq 0$，则从推论(8.1.19)可得出结论，如果 $\rho(A)=0$，则把推论(8.1.20)应用于 B 可得出结论.

因为我们马上就会有关于非负矩阵的谱半径的较好上界，定理(8.18)将用来求任意矩阵的谱半径的上界.

8.1.21 引理 设 $A\in M_n$，且假定 $A\geqslant 0$. 如果 A 的各个行和是常数，则 $\rho(A)=\|A\|_\infty$. 如果 A 的各个列和是常数，则 $\rho(A)=\|A\|_1$.

证明：我们知道，$\rho(A)\leqslant\|A\|$ 对任意矩阵范数 $\|\cdot\|$ 成立. 但是，如果各行和是常数，则 $x=[1, \cdots, 1]^T$ 是关于特征值 $\|A\|_\infty$ 的特征向量. 把同样的论证应用于 A^T 便得到关于列和的论断. □

8.1.22 定理 设 $A\in M_n$，且假定 $A\geqslant 0$. 则

$$\min_{1\leqslant i\leqslant n}\sum_{j=1}^{n}a_{ij}\leqslant\rho(A)\leqslant\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}a_{ij} \tag{8.1.23}$$

且

$$\min_{1\leqslant j\leqslant n}\sum_{i=1}^{n}a_{ij}\leqslant\rho(A)\leqslant\max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}a_{ij} \tag{8.1.24}$$

证明：设 $\alpha=\min\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^{n}a_{ij}$，然后构造新矩阵 B，使 $A\geqslant B\geqslant 0$，且 $\sum\limits_{j=1}^{n}b_{ij}\equiv\alpha$ 对所有 $i=1,2,\cdots$ 成立. 例如，如果 $\alpha=0$，令 $B=0$. 又如果 $\alpha>0$，可以令 $b_{ij}=\alpha a_{ij}(\sum\limits_{j=1}^{n}a_{ij})^{-1}$. 根据引理(8.1.21)，$\rho(B)=\alpha$，又根据推论(8.1.19)，$\rho(B)\leqslant\rho(A)$. 相应的上界易用类似的方法来证明. 把行和界应用于 A^T 便得到列和界. □

练习 证明上述结果中关于上界的论断.

8.1.25 推论 设 $A\in M_n$. 如果 $A\geqslant 0$，且对所有 $i=1,2,\cdots,n$ 有 $\sum\limits_{j=1}^{n}a_{ij}>0$，则 $\rho(A)>0$. 特别是，如果 $A>0$，或者如果 A 是不可约非负矩阵，则 $\rho(A)>0$. 492

练习 证明不可约矩阵不可能有零行或零列.

因为只要 S 可逆就有 $\rho(S^{-1}AS)=\rho(A)$，所以可以引进某些自由参数来推广上述定理. 如果 $S=\mathrm{diag}(x_1,\cdots,x_n)$，且所有 $x_i>0$，则当 $A\geqslant 0$ 时 $S^{-1}AS\geqslant 0$. 把定理(8.1.22)应用于 $S^{-1}AS=(a_{ij}x_jx_i^{-1})$，便得到下述更一般的结果.

8.1.26 定理 设 $A\in M_n$，且假定 $A\geqslant 0$. 则对任意正向量 $x\in\mathbf{C}^n$，有

$$\min_{1\leqslant i\leqslant n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j\leqslant\rho(A)\leqslant\max_{1\leqslant i\leqslant n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j \tag{8.1.27}$$

和

$$\min_{1\leqslant j\leqslant n}x_j\sum_{i=1}^{n}\frac{a_{ij}}{x_i}\leqslant\rho(A)\leqslant\max_{1\leqslant j\leqslant n}x_j\sum_{i=1}^{n}\frac{a_{ij}}{x_i}. \tag{8.1.28}$$

8.1.29 推论 设 $A\in M_n$，$x\in\mathbf{R}^n$，且假定 $A\geqslant 0$ 和 $x>0$. 如果 α、$\beta\geqslant 0$ 使 $\alpha x\leqslant Ax\leqslant\beta x$，则 $\alpha\leqslant\rho(A)\leqslant\beta$. 如果 $\alpha x<Ax$，则 $\alpha<\rho(A)$；如果 $Ax<\beta x$，则 $\rho(A)<\beta$.

证明：如果 $\alpha x\leqslant Ax$，则 $\alpha\leqslant\min\limits_{1\leqslant i\leqslant n}x_i^{-1}\sum\limits_{j=1}^{n}a_{ij}x_j$. 由定理(8.1.26)得知 $\alpha\leqslant\rho(A)$. 如果 $\alpha x<Ax$，则有某个 $\alpha'>\alpha$，使得 $\alpha'x\leqslant Ax$. 在这种情形 $\rho(A)\geqslant\alpha'>\alpha$，所以 $\rho(A)>\alpha$. 类似地可证明上界. □

练习 完成推论(8.1.29)的证明.

8.1.30 推论 设 $A\in M_n$，且假定 A 非负. 如果 A 有正特征向量，则相应的特征值是 $\rho(A)$；即如果 $Ax=\lambda x$ 且 $x>0$ 和 $A\geqslant 0$，则 $\lambda=\rho(A)$.

证明：如果 $x>0$ 且 $Ax=\lambda x$，则 $\lambda\geqslant 0$ 且 $\lambda x\leqslant Ax\leqslant\lambda x$. 另一方面，由推论(8.1.29)可知，$\lambda\leqslant\rho(A)\leqslant\lambda$. □

8.1.31 推论 设 $A\in M_n$，且假定 A 是非负矩阵. 如果 A 有正特征向量，则

$$\rho(A)=\max_{x>0}\min_{1\leqslant i\leqslant n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j=\min_{x>0}\max_{1\leqslant i\leqslant n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j. \tag{8.1.32}$$ 493

练习 证明上述推论. **提示**：在(8.1.27)中采用正特征向量.

8.1.33 推论 设 $A\in M_n$，且假定 A 是非负矩阵. 如果 A 有正特征向量 x，则对所有 $m=1$,

2. …和所有 $i=1, \cdots, n$，有

$$\sum_{j=1}^{n} a_{ij}^{(m)} \leqslant \left[\frac{\max\limits_{1\leqslant k\leqslant n} x_k}{\min\limits_{1\leqslant k\leqslant n} x_k}\right]\rho(A)^m \quad 和 \quad \left[\frac{\min\limits_{1\leqslant k\leqslant n} x_k}{\max\limits_{1\leqslant k\leqslant n} x_k}\right]\rho(A)^m \leqslant \sum_{j=1}^{n} a_{ij}^{(m)}, \tag{8.1.34}$$

其中 $A^m \equiv [a_{ij}^{(m)}]$. 特别是，如果 $\rho(A)>0$，则对 $m=1, 2, \cdots$，$[\rho(A)^{-1}A]^m$ 各元一致有界.

证明：如果 $Ax=\rho(A)x$，则 $A^m x=\rho(A)^m x$. 如果 $A\geqslant 0$，则 $A^m\geqslant 0$，且有

$$\rho(A)^m \max_{1\leqslant k\leqslant n} x_k \geqslant \rho(A^m)x_i = [A^m x]_i = \sum_{j=1}^{n} a_{ij}^{(m)} x_j$$

$$\geqslant \left(\min_{1\leqslant k\leqslant n} x_k\right) \sum_{j=1}^{n} a_{ij}^{(m)}$$

对任意 $i=1, 2, \cdots, n$ 成立. 因为 $x>0$，所确定的上界可由除法得出. 类似地有

$$\rho(A)^m \min_{1\leqslant k\leqslant n} x_k \leqslant \rho(A)^m x_i = [A^m x]_i = \sum_{j=1}^{n} a_{ij}^{(m)} x_j$$

$$\leqslant \left(\max_{1\leqslant k\leqslant n} x_k\right) \sum_{j=1}^{n} a_{ij}^{(m)}$$

对任意 $i=1, \cdots, n$ 成立. 因为 $x>0$，所确定的下界可由除法得出. □

习题

1. 如果 $A\geqslant 0$，且对某个 k，$A^k>0$，证明 $\rho(A)>0$.

2. 给出一个 2×2 矩阵 A，使得 $A\geqslant 0$，A 不是正的，且 $A^2>0$.

3. 假定 $A\geqslant 0$，且 $A\neq 0$. 如果 A 有正特征向量，证明 $\rho(A)>0$.

4. 如果 $\rho(A)<1$，我们知道，当 $m\to\infty$ 时 $A^m\to 0$. 如果 $A\geqslant 0$，且 A 有正特征向量，试用推论(8.1.33)证明，$|A^m|\leqslant\rho(A)^m C(A)$ 对所有 $m=1, 2, \cdots$ 成立，其中 $C(A)$ 是常数矩阵. 试考察 $A=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$，说明 A 有正特征向量的假定不能省略. 试从推论(5.6.13)来讨论这两个结果.

[494]

5. 如果 $A\geqslant 0$ 有特征向量，证明 A 相似于其各行和为常数的非负矩阵. 这个常数是什么? **提示**：利用定理(8.1.26)前面的注释.

6. 我们将在(8.4)节证明，非负不可约矩阵必定有正特征向量. 说明非负矩阵可以有正特征向量而且是可约的.

7. 设 $A=[a_{ij}]\in M_n$ 是非负矩阵且有正特征向量 $x=[x_i]$. 试用(8.1.33)证明，

(a) $\frac{1}{n}\rho(A)^m\left[\frac{\min\limits_{1\leqslant k\leqslant n} x_k}{\max\limits_{1\leqslant k\leqslant n} x_k}\right]\leqslant \max\limits_{1\leqslant p\leqslant n} a_{ip}^{(m)}$ 对每个 $i=1, 2, \cdots, n$ 成立；

(b) $\lim\limits_{m\to\infty}\left[\sum_{p=1}^{n} a_{ip}^{(m)}\right]^{1/m}=\rho(A)$ 对每个 $i=1, 2, \cdots, n$ 成立.

对所有 $m=1, 2, \cdots$，记 $A^m=[a_{ij}^{(m)}]$.

8.2 正矩阵

对正矩阵来说，非负矩阵理论显示出它的最简单和最优美的形式，关于这种情形，O. Perron 在 1607 年做出了主要的贡献.

8.2.1 引理 设 $A\in M_n$，且假定 $A>0$，$Ax=\lambda x$，$x\neq 0$，以及 $|\lambda|=\rho(A)$. 则 $A|x|=\rho(A)|x|$ 且 $|x|>0$.

证明：计算

$$\rho(A)|x|=|\lambda||x|=|\lambda x|=|Ax|\leqslant |A||x|=A|x|$$

因而 $y\equiv A|x|-\rho(A)|x|\geqslant 0$，因为 $|x|\geqslant 0$ 且 $|x|\neq 0$，所以由(8.1.14)可知 $A|x|>0$. 又推论(8.1.25)保证 $\rho(A)>0$. 所以，当 $y=0$ 时有 $A|x|=\rho(A)|x|$ 和 $|x|=\rho(A)^{-1}A|x|>0$. 如果 $y\neq 0$，令 $z\equiv A|x|>0$，再运用(8.1.14)：

$$0<Ay=Az-\rho(A)z \quad 或 \quad Az>\rho(A)z$$

但是根据推论(8.1.29)，我们有荒谬的结论 $\rho(A)>\rho(A)$. 从而得知 $y=0$，证毕. □

由这个技术性引理，容易推导出关于正矩阵的第一个基本结果. [495]

8.2.2 定理 设 $A\in M_n$，且假定 A 是正矩阵. 则 $\rho(A)>0$，$\rho(A)$是 A 的特征值，且存在正向量 x 使得 $Ax=\rho(A)x$.

证明：存在特征值 λ，且 $|\lambda|=\rho(A)>0$，而相应的特征向量 $x\neq 0$. 根据引理，所要求的向量是 $|x|$. □

练习 如果 $A\in M_n$，且 $A>0$，试用推论(8.1.31)证明

$$\rho(A)=\max_{x>0}\min_{1\leqslant i\leqslant n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j=\min_{x>0}\max_{1\leqslant i\leqslant n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j.$$

只要稍微强化一下引理(8.2.1)的叙述，就可以加深对 A 的特征值的估计的认识.

8.2.3 引理 设 $A\in M_n$，且假定 $A>0$，$Ax=\lambda x$，$x\neq 0$，以及 $|\lambda|=\rho(A)$. 则对某个 $\theta\in\mathbf{R}$，$e^{-i\theta}x=|x|>0$.

证明：假设条件保证 $|Ax|=|\lambda x|=\rho(A)|x|$，且由引理(8.2.1)可知，$A|x|=\rho(A)|x|$ 且 $|x|>0$. 综合这两个恒等式和三角不等式，对每个 $k=1,\cdots,n$，有

$$\rho(A)|x_k|=|\lambda||x_k|=|\lambda x_k|=\left|\sum_{p=1}^{n}a_{kp}x_p\right|$$

$$\leqslant\sum_{p=1}^{n}|a_{kp}||x_p|=\sum_{p=1}^{n}a_{kp}|x_p|=\rho(A)|x_k|.$$

因此，在三角不等式中等式必须成立，因而(非零)复数 $a_{kp}x_p$，$p=1,\cdots,n$，一定都位于复平面的同一条射线上. 如果用 θ 表示它们的公共辐角，则 $e^{-i\theta}a_{kp}x_p>0$ 对所有 $p=1,\cdots,n$ 成立. 但是，因为所有 $a_{kp}>0$，所以有 $e^{-i\theta}x>0$. □

8.2.4 定理 设 $A\in M_n$，且假定 A 是正矩阵. 则对每个特征值 $\lambda\neq\rho(A)$有 $|\lambda|<\rho(A)$.

证明：根据定义，$|\lambda|\leqslant\rho(A)$对 A 的所有特征值 λ 成立. 假如 $|\lambda|=\rho(A)$，且 $Ax=$

λx，$x\neq 0$. 根据引理(8.2.3)，对某个 $\theta\in\mathbf{R}$ 有 $w=e^{-i\theta}x>0$，所以 $Aw=\lambda w$. 但是根据推论(8.1.30)有 $\lambda=\rho(A)$. □

现在我们知道，如果 $A>0$，则 $\rho(A)$ 可以看成严格最大模特征值，而不会有其他可能性. 下一个结果说明，$\rho(A)$ 是几何重数为 1 的特征值；即相应于 $\rho(A)$ 的特征空间有维数 1. 实际上，我们马上就会看到，$\rho(A)$ 的代数重数也是 1. [496]

8.2.5 定理 设 $A\in M_n$，且假定 $A>0$，w 和 z 是适 $Aw=\rho(A)w$ 和 $Az=\rho(A)z$ 的非零向量. 则存在某个 $\alpha\in\mathbf{C}$ 使得 $w=\alpha z$.

证明：根据引理(8.2.3)，存在实数 θ_1 和 θ_2 使得 $p\equiv e^{-i\theta_1}z>0$ 和 $q\equiv e^{-i\theta_2}w>0$. 令 $\beta\equiv\min\limits_{1\leq i\leq n} q_ip_i^{-1}$ 且定义 $r\equiv q-\beta p$. 注意到 $r\geq 0$ 且 r 至少有一个坐标为 0，所以 r 不是正向量. 但是 $Ar=Aq-\beta Ap=\rho(A)q-\beta\rho(A)p=\rho(A)r$，所以，如果 $r\neq 0$，由(8.1.14)可知，$r=\rho(A)^{-1}Ar>0$. 因为这不成立，得出 $r=0$，因而 $q=\beta p$ 且 $w=\beta e^{i(\theta_2-\theta_1)}z$. □

8.2.6 推论 设 $A\in M_n$，且假定 $A>0$. 则存在唯一向量 x，使得 $Ax=\rho(A)x$，$x>0$ 且 $\sum\limits_{i=1}^{n}x_i=1$.

练习 证明推论(8.2.6).

推论(8.2.6)中所描述的唯一的正规化特征向量常常称为 A 的 Perron 向量；$\rho(A)$ 常常称为 A 的 Perron 根. 当然，如果 A 是正矩阵，则 A^T 亦是，所以，上述所有结果也可以应用于 A^T. A^T 的 Perron 向量称为 A 的左 Perron 向量.

练习 如果 $A\in M_n$，且 $A>0$，又如果存在某个 $x\in\mathbf{C}^n$，使得 $x\geq 0$，$x\neq 0$，且 $Ax=\lambda x$，证明 x 是 A 的 Perron 向量的倍数且 $\lambda=\rho(A)$.

我们对研究 $m\to\infty$ 时幂 A^m 的变化过程感兴趣，因为这些幂出现在数值分析的应用中以及概率论中的 Markov 链理论的应用中. 下面的引理把对于非负矩阵的各个极限定理是本质的诸条件分离开来. 注意，如果 $A>0$ 且 $\lambda=\rho(A)$，则所有假设条件都适合.

[497] **8.2.7 引理** 设 $A\in M_n$ 是给定的矩阵，$\lambda\in\mathbf{C}$ 是给定的数，且假定 x 和 y 是适合

(1) $Ax=\lambda x$；

(2) $A^Ty=\lambda y$；

(3) $x^Ty=1$；

的向量. 定义 $L\equiv xy^T$. 则

(a) $Lx=x$ 且 $y^TL=y^T$；

(b) $L^m=L$ 对所有 $m=1,2,\cdots$ 成立；

(c) $A^mL=LA^m=\lambda^mL$ 对所有 $m=1,2,\cdots$ 成立；

(d) $L(A-\lambda L)=0$；

(e) $(A-\lambda L)^m=A^m-\lambda^mL$ 对所有 $m=1,2,\cdots$ 成立；

(f) $A-\lambda L$ 的每个非零特征值也是 A 的特征值.

如果另外还假定

(4) $\lambda \neq 0$

(5) λ 是的 A 的特征值，且有几何重数 1；于是还有

(g) λ 不是 $A-\lambda L$ 的特征值；即 $\lambda I-(A-\lambda L)$ 是可逆矩阵.

最后，如果假定

(6) $|\lambda|=\rho(A)>0$；

(7) λ 是 A 的模为 $\rho(A)$ 的仅有特征值；又如果把 A 的特征值排成顺序 $|\lambda_1|\leqslant|\lambda_2|\leqslant\cdots\leqslant|\lambda_{n-1}|<|\lambda_n|=|\lambda|=\rho(A)$，则

(h) $\rho(A-\lambda L)\leqslant|\lambda_{n-1}|<\rho(A)$；

(i) 当 $m\to\infty$ 时，$(\lambda^{-1}A)^m-L=(\lambda^{-1}A-L)^m\to L$；

(j) 对每个适合 $[|\lambda_{n-1}|/\rho(A)]<r<1$ 的 r，存在某个 $C=C(r,A)$，使得 $\|(\lambda^{-1}A)^m-L\|_\infty<Cr^m$ 对所有 $m=1,2,\cdots$成立.

证明：论断(a)，(b)和(c)可直接从假定(1)，(2)和(3)推出；注意(3)推出，x 和 y 都是非零向量. 论断(d)可以由(b)和(c)推出. 命题(e)可以用(b)和(c)来归纳证明. 如果 $\mu\neq 0$ 是 $A-\lambda L$ 的特征值，又如果对某个 $w\neq 0$ 有 $(A-\lambda L)w=\mu w$，则 $L(A-\lambda L)w=0w=0=\mu Lw$，因而 $Lw=0$. 因此，$(A-\lambda L)w=Aw=\mu w$，所以 μ 也是 A 的特征值，(f)得证.

498

如果现在引用假设(4)且取 $\mu=\lambda$，上述论证说明，如果 w 是 $A-\lambda L$ 的特征向量，则 w 也是 A 的 λ 的特征向量. 根据假设(5)，我们一定可得出 $w=\alpha x$ 对某个 $\alpha\neq 0$ 成立. 另一方面，$\mu w-\lambda w=(A-\lambda L)w=(A-\lambda L)\alpha x=\alpha\lambda x-\lambda\alpha x=0$，又因为 $\lambda\neq 0$ 且 $w\neq 0$，所以这是不可能的. 这个矛盾证明了(g). 因为(f)，我们知道，或者对 A 的某个特征值 λ_k 有 $\rho(A-\lambda L)=|\lambda_k|$，或者 $\rho(A-\lambda L)=0$. 因为已给出了 A 的特征值的递增顺序，且 $|\lambda_n|=|\lambda|=\rho(A)$，所以由(g)可知，在这两种情形，$\rho(A-\lambda L)\leqslant|\lambda_{n-1}|$. 因此(h)中的不等式可直接由(7)推出. 把(h)和(e)结合起来不难算出，当 $m\to\infty$ 时 $(\lambda^{-1}A-L)^m=(\lambda^{-1}A)^m-L\to 0$，这是因为 $\rho(\lambda^{-1}A-L)=\rho(A-\lambda L)/\rho(A)\leqslant|\lambda_{n-1}|/\rho(A)\leqslant 1$. (j)中的收敛速度是推论(5.6.13)应用于矩阵 $\lambda^{-1}A-L$ 的直接推论，其中选取 ϵ 适合 $\rho(\lambda^{-1}A-L)+\epsilon\leqslant[|\lambda_{n-1}|/\rho(A)]+\epsilon<r<1$. □

练习 给出引理中(a)，(b)和(c)的证明细节.

8.2.8 定理 设 $A\in M_n$ 且假定 $A>0$. 则

$$\lim_{m\to\infty}[\rho(A)^{-1}A]^m=L,$$

其中，$L\equiv xy^T$，$Ax=\rho(A)x$，$A^Ty=\rho(A)y$，$x>0$，$y>0$ 且 $x^Ty=1$.

证明：引理的假定(1)—(7)为 $\lambda=\rho(A)$ 所满足，x 是 A 的 Perron 向量，且 $y=(x^Tz)^{-1}z$，其中 z 是 A^T 的 Perron 向量. 结论可由(i)推出. □

8.2.9 推论 如果 $A\in M_n$ 且 $A>0$，则 $L\equiv\lim_{m\to\infty}[\rho(A)^{-1}A]^m$ 是秩 1 的正矩阵.

8.2.10 定理 如果 $A\in M_n$ 且 $A>0$. 则 $\rho(A)$ 是代数重数为 1 的特征值；即 $\rho(A)$ 是特征方程 $p_A(t)=0$ 的单根.

证明：根据 Schur 三角化定理(2.3.1)，可以记 $A=U\Delta U^*$，其中，U 是酉矩阵，Δ 是主对角元为 $\rho,\cdots,\rho,\lambda_{k+1},\cdots,\lambda_n$ 的上三角矩阵，且 $\rho=\rho(A)$ 是代数重数 $k\geqslant 1$ 的特征值；对所有

[499] $i=k+1, \cdots, n$，特征值的模都严格小于 $\rho(A)$. 另一方面，

$$L=\lim_{m\to\infty}[\rho(A)^{-1}A]^m$$

$$=U\lim_{m\to\infty}\begin{bmatrix}1 & & & & & * \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \frac{\lambda_{k+1}}{\rho} & & \\ & & & & \ddots & \\ & 0 & & & & \frac{\lambda_n}{\rho}\end{bmatrix}^m U^*$$

$$=U\begin{bmatrix}1 & & & & & * \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0\end{bmatrix}U^*,$$

其中，在后两个表示中对角元 1 重复 k 次，在最后一个表示式中对角 0 元重复 $n-k$ 次. 因为最后表示式中的上三角矩阵至少有秩 k，又因为 L 有秩 1，得知，$k>1$ 是不可能的. □

我们现在总结一下在本节中得到的关于正矩阵的基本结果.

8.2.11 Perron 定理　如果 $A\in M_n$ 且 $A>0$，则

(a) $\rho(A)>0$；

(b) $\rho(A)$是 A 的特征值；

(c) 存在 $x\in \mathbf{C}^n$，适合 $x>0$ 且 $Ax=\rho(A)x$；

(d) $\rho(A)$是 A 的代数(因而几何)单重特征值；

(e) $|\lambda|<\rho(A)$对每个特征值 $\lambda\neq\rho(A)$成立，即 $\rho(A)$是唯一的最大模特征值；

(f) 当 $m\to\infty$时$[\rho(A)^{-1}A]^m\to L$，其中 $L\equiv xy^T$，$Ax=\rho(A)x$，$A^Ty=\rho(A)y$，$x>0$，$y>0$ 且 $x^Ty=1$.

Perron 定理有许多应用. 一个优美而又有效的应用是，利用占优非负矩阵的谱半径和主 [500] 对角元得到了矩阵 A 的特征值包含区域.

8.2.12 定理(樊畿)　设 $A=[a_{ij}]\in M_n$，且假定 $B=[b_{ij}]\in M_n$ 有非负元和 $B\geqslant|A|$. 则 A 的每个特征值位于区域

$$\bigcup_{i=1}^{n}\{z\in\mathbf{C}:|z-a_{ii}|\leqslant\rho(B)-b_{ii}\}$$

中.

证明：可以假定 $B>0$，因为如果 B 的某个元是零，我们可以考虑 $B_\varepsilon\equiv[b_{ij}+\varepsilon]$，其中 $\varepsilon>0$；$B_\varepsilon>|A|$，且当 $\varepsilon\to 0$ 时 $\rho(B_\varepsilon)-(b_{ii}+\varepsilon)\to\rho(B)-b_{ii}$. 根据 Perron 定理，存在正向量 x 使

得 $Bx=\rho(B)x$，因而对所有 $i=1, 2, \cdots, n$ 有

$$\sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| x_j \leqslant \sum_{\substack{j=1\\ j\neq i}}^{n} b_{ij}x_j = \rho(B)x_i - b_{ii}x_i.$$

因此，对所有 $i=1, 2, \cdots, n$ 有

$$\frac{1}{x_i}\sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| x_j \leqslant \rho(B) - b_{ii}.$$

在推论(6.1.6)中取 $p_i=x_i$ 便推出要求的结果. □

(8.2.11)中的(f)保证确定的极限存在，而(8.2.7)中的(j)给出了关于收敛速度的一个上界

$$\| [\rho(A)^{-1}A]^m - L \|_\infty < C r^m,$$

这个上界对某个与 A 和 r 有关的正常数 C 和任何适合

$$\frac{|\lambda_{n-1}|}{\rho(A)} < r < 1$$

的 r 成立，其中 λ_{n-1} 是 A 的有第二大模的特征值. 即使 $\rho(A)$ 已知或容易估计，但是要想得出关于比值 $|\lambda_{n-1}|/\rho(A)$ 的有用界，计算或估计 $|\lambda_{n-1}|$ 也许不方便或者不可能. 在这种情形，知道一个容易计算的界可能很有用，这个界属于 E. Hopf，它对任何正矩阵 $A=[a_{ij}]\in M_n$ 都成立：

$$\frac{|\lambda_{n-1}|}{\rho(A)} \leqslant \frac{M-\mu}{M+\mu} < 1,$$

其中 $M=\max\{a_{ij}: i, j=1, 2, \cdots, n\}$ 而 $\mu=\min\{a_{ij}: i, j=1, 2, \cdots, n\}$.

习题

1. 如果 $A>0$，x 是 A 的 Perron 向量，又如果 z 是 A^T 的 Perron 向量，证明 $x^Tz>0$.　501

2. 如果 $\Delta\in M_n$ 是具有 k 个非零主对角元的上三角矩阵，证明 $\operatorname{rank}\Delta\geqslant k$. 用例子说明，在这些条件下，有可能秩大于 k.

3. 把本节推导出的结果应用于矩阵

$$A=\begin{bmatrix}1-\alpha & \beta\\ \alpha & 1-\beta\end{bmatrix},\quad 0<\alpha,\ \beta<1,$$

且与(8.0)节中的相关结论相比较.

4. 如(8.0)节中所描述的那样，考虑具有 $n>2$ 个城市的一般城市间的人口流动模型. 如果所有人口流动系数 a_{ij} 是正的，那么，当 $m\to\infty$ 时，人口分布 $p^{(m)}$ 的渐近变化过程是什么？

5. 如果 $A>0$，详细描述当 $m\to\infty$ 时 A^m 的渐近变化过程. **提示**：存在三种情形：$A^m\to 0$，A^m 发散和收敛于正矩阵. 说明并分析每种情形.

6. 在推论(6.1.8)下面有一个讨论 2×2 正矩阵的练习试用定理(8.2.2)下面的练习讨论这个例子.

7. 设 $A, B\in M_n$，且假定 $A>B>0$. 试且 $\rho(B)$ 的"min max"特征证明 $\rho(A)>\rho(B)$. **提示**：设 x 是 A 的适合 $Ax>Bx$ 的 Perron 向量.

8. 如果 $A>0$，且 x 是 A 的 Perron 向量，证明

$$\rho(A)=\sum_{i,j=1}^{n}a_{ij}x_j.$$

由定义可知，$x_1+\cdots+x_n=1$.

9. 如果正矩阵是非奇异的，证明其逆矩阵可能不是非负的. 如果非负矩阵 A 是非奇异的，证明，仅当 A 的每一列中恰好有一个非零元时，其逆矩阵才可以是非负的. 这样的矩阵与置换矩阵有何关系?

10. 试给出定理(8.2.10)的下述另一个证明的细节：如果 $\rho=\rho(A)$ 有代数重数 $k>1$. 又如果 y 和 x 分别是 A 的左和右 Perron 向量，则存在非零向量 z 使得 $x=(A-\rho I)z$. 另一方面，$y^Tx=y^T(A-\rho I)z=0^Tz=0$，因为 $y^Tx>0$，所以这是不可能的. **提示**：因为 ρ 的几何重数是 1，所以 $A-\rho I$ 的 Jordan 形一定恰好有一个幂零子块，它至少应该是二阶. 证明秩为 $k-1(k>1)$ 的任一幂零矩阵 $B\in M_k$ 有以下性质：如果对某个 $u\in\mathbf{C}^k$ 有 $Bu=0$，则存在某个 $v\in\mathbf{C}^k$ 使得 $Bv=u$.

502

进一步阅读 集中论述各种界(其中包括本节最后一段提到的 Hopf 界)以及有关这方面的众多文献，可参看 U. Rothblum and C. Tan, "Upper Bounds on the Maximum Modulus of Subdominant Eigenvalues of Nonnegative Matrices," *Linear Algebra Appl*. 66(1985), 45-86. 也可参看[Kel]chapter Ⅱ, theorem 2.

8.3 非负矩阵

因为在实际中常常遇到不是正矩阵的非负矩阵，为此，需要考虑把上一节中所论述的理论推广到矩阵的元素不都是严格正的情形. 我们可能希望，这种推广可通过取适当的极限来完成，这对于某些结果是可以办到的. 但是，令人失望的是，像秩与维数这样一些量不是连续函数，所以极限论证只是部分有效. 在 Perron 定理中可以通过取极限作推广的仅有结果包括在下面的定理中.

8.3.1 定理 如果 $A\in M_n$，且 $A\geqslant 0$，则 $\rho(A)$ 是 A 的特征值，且存在非负向量 $x\geqslant 0$，$x\neq 0$，使得 $Ax=\rho(A)x$.

证明：对任意 $\varepsilon>0$，定义 $A(\varepsilon)\equiv[a_{ij}+\varepsilon]>0$. 用 $x(\varepsilon)$ 表示 $A(\varepsilon)$ 的 Perron 向量，所以 $x(\varepsilon)>0$ 且 $\sum_{i=1}^{n}x(\varepsilon)_i=1$. 因为向量集$\{x(\varepsilon):\varepsilon>0\}$包含在紧集$\{x: x\in\mathbf{C}^n, \|x\|_1\leqslant 1\}$中，所以存在单调递减序列 ε_1, ε_2, …且 $\lim_{k\to\infty}\varepsilon_k=0$，使得 $\lim_{k\to\infty}x(\varepsilon_k)\equiv x$ 存在. 因为对所有 $k=1, 2, \cdots$ 有 $x(\varepsilon_k)>0$，所以一定有 $x=\lim_{k\to\infty}x(\varepsilon_k)\geqslant 0$；又

$$\sum_{i=1}^{n}x_i=\lim_{k\to\infty}\sum_{i=1}^{n}x(\varepsilon_k)_i=1$$

所以 $x=0$ 是不可能的. 根据定理(8.1.18)，$\rho(A(\varepsilon_k))\geqslant\rho(A(\varepsilon_{k+1}))\geqslant\cdots\geqslant\rho(A)$ 对所有 $k=1, 2, \cdots$ 成立，所以实数序列$\{\rho(A(\varepsilon_k))\}_{k=1,2,\cdots}$是单调递减序列. 因此，$\rho\equiv\lim_{k\to\infty}\rho(A(\varepsilon_k))$存在且 $\rho\geqslant\rho(A)$. 但是由事实

$$Ax=\lim_{k\to\infty}A(\varepsilon_k)x(\varepsilon_k)=\lim_{k\to\infty}\rho(A(\varepsilon_k))x(\varepsilon_k)$$

$$= \lim_{k\to\infty} \rho(A(\varepsilon_k)) \lim_{k\to\infty} x(\varepsilon_k) = \rho x$$

503

和事实 $x\neq 0$，推出 ρ 是 A 的特征值. 另一方面，$\rho\leqslant\rho(A)$. 因此一定有 $\rho=\rho(A)$. □

关于谱半径的变分特征部分(8.1.31)可以推广到一般非负矩阵和非负向量，不过其证明是颇不相同的.

8.3.2 定理 设 $A\in M_n$，$A\geqslant 0$，$x\in\mathbf{C}^n$，$x\geqslant 0$ 且 $x\neq 0$. 如果对某个 $\alpha\in\mathbf{R}$，$Ax\geqslant\alpha x$，则 $\rho(A)\geqslant\alpha$.

证明：设 $A=[a_{ij}]$，设 $\varepsilon>0$，且定义 $A(\varepsilon)=[a_{ij}+\varepsilon]$. 于是 $A(\varepsilon)>0$，所以有正的左 Perron 向量 $y(\varepsilon)$；即 $y(\varepsilon)^TA(\varepsilon)=\rho(A(\varepsilon))y(\varepsilon)^T$. 已知 $Ax-\alpha x\geqslant 0$，所以 $A(\varepsilon)x-\alpha x>Ax-\alpha x\geqslant 0$，因而 $y(\varepsilon)^T[A(\varepsilon)x-\alpha x]=[\rho(A(\varepsilon))-\alpha]y(\varepsilon)^Tx\geqslant 0$. 因为 $y(\varepsilon)^Tx>0$，对所有 $\varepsilon>0$，有 $\rho(A(\varepsilon))-\alpha\geqslant 0$. 但当 $\varepsilon\to 0$ 时 $\rho(A(\varepsilon))\to\rho(A)$，所以 $\rho(A)\geqslant\alpha$. □

8.3.3 推论 如果 $A\in M_n$，且 $A\geqslant 0$，则

$$\rho(A) = \max_{\substack{x\geqslant 0\\ x\neq 0}} \min_{\substack{1\leqslant i\leqslant n\\ x_i\neq 0}} \frac{1}{x_i}\sum_{j=1}^{n} a_{ij}x_j.$$

证明：如果 $A\geqslant 0$，$x\geqslant 0$，且 $x\neq 0$，又如果选取

$$\alpha \equiv \min_{x_i\neq 0}\sum_{j=1}^{n}\frac{a_{ij}x_j}{x_i},$$

则根据定理(8.3.2)，$Ax\geqslant\alpha x$，因而 $\alpha\leqslant\rho(A)$. 但是如果选取 x 为由定理(8.3.1)保证其存在的特征向量，那么我们看到这个上界可以用 $\alpha=\rho(A)$ 来达到. □

练习 考虑 $A=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}$ 和 $x=\begin{bmatrix}1\\ 0\end{bmatrix}$，证明，如果 x 不是正向量，(8.1.29)的上界不一定成立. 证明(8.1.32)的"min max"特征一般也不成立. 然而，如上述推论所证明的那样，"max min"特征的确可以推广.

添加一个假设条件，可以稍微强化定理(8.3.2)以给出关于向量 x 的某些信息.

8.3.4 定理 设 $A\in M_n$，且 $A\geqslant 0$，再假定 A 有正左特征向量. 如果 $x\geqslant 0$，$x\neq 0$，又如果 $Ax\geqslant\rho(A)x$，则 $Ax=\rho(A)x$.

504

证明：设 $y>0$ 适合 $A^Ty=\rho(A)y$，且假定 $x\geqslant 0$ 适合 $x\neq 0$ 和 $Ax-\rho(A)x\geqslant 0$. 则

$$y^T[Ax-\rho(A)x] = \rho(A)y^Tx - \rho(A)y^Tx = 0,$$

所以一定有 $Ax-\rho(A)x=0$. □

不添加假设条件，不可能比定理(8.3.1)更进一步把 Perron 定理(8.2.11)推广到非负矩阵.

当 $A\in M_n$ 且 $A\geqslant 0$ 时，非负特征值 $\rho(A)$ 称为 A 的 Perron 根. 因为相应于非负矩阵的 Perron 根的特征向量未必是唯一确定的，所以对于一般非负矩阵(不同于 A 是正矩阵的情形)，没有关于"Perron 向量"的完全确定的概念. 例如，非负矩阵 $A=I$ 有每个非负向量作为相应于 Perron 根 $\rho(A)=1$ 的特征向量.

习题

1. 用例子说明，在 Perron 定理(8.2.11)中的不包括在定理(8.3.1)中的各项命题对所有非

负矩阵一般不成立. **提示**: 考虑$\begin{bmatrix}0&1\\0&0\end{bmatrix}$, $\begin{bmatrix}1&0\\0&1\end{bmatrix}$和$\begin{bmatrix}0&1\\1&0\end{bmatrix}$.

2. 如果 $A\geqslant 0$, 且对某个 $k\geqslant 1$ 有 $A^k>0$, 证明 A 有正特征向量.

3. 如果 $A\geqslant 0$ 有非负特征向量 x, 其中有 $r\geqslant 1$ 个正分量和 $n-r$ 个零分量, 证明, A 经置换相似可以变成形式$\begin{bmatrix}B&C\\0&D\end{bmatrix}$, 其中 $B\in M_r$, $C\in M_{r,(n-r)}$, $D\in M_{n-r}$; B, C 和 D 是非负矩阵, 且 B 有正特征向量. 如果 $r<n$, 证明 A 一定是可约矩阵.

4. 用例子说明, 推论(8.1.30)的下述推广不成立: 设 $A\geqslant 0$. 如果 A 有非负特征向量 $x\geqslant 0$, $x\neq 0$, 则 $Ax=\rho(A)x$.

5. 考虑矩阵 $A=\begin{bmatrix}0&1\\0&1\end{bmatrix}$和向量 $x=[1,\ 2]^T$. 说明, 如果只假定 $A\geqslant 0$, 定理(8.3.4)是不成立的. A 的左和右 Perron 向量是什么?

6. 如果 $A\geqslant 0$, 证明, 存在与 A 交换的正矩阵 B, 当且仅当 A 有全是正的左和右特征向量. **提示**: 如果 x 和 y 是 A 的正右特征向量和正左特征向量, 设 $B=xy^T$. 反之, 如果 $x\geqslant 0$, 且 $Ax=\rho(A)x$, 考虑 $BAx=ABx=B\rho(A)x>0$.

505

7. 如果 $A=[a_{ij}]\in M_n$ 是非负三对角矩阵, 证明 A 的所有特征值都是实的. **提示**: 首先证明, 如果所有下和上对角元都是正的, 则可以求得正对角矩阵 D 使得 $D^{-1}AD$ 是对称矩阵. 然后证明对角线以上或以下的 0 元是无关紧要的.

8. 设非负矩阵 $A\in M_n$ 是给定的. 证明, 或者 A 是不可约的, 或者存在置换矩阵 P 使得

$$P^TAP=\begin{bmatrix}A_1 & & *\\ & \ddots & \\ 0 & & A_k\end{bmatrix},$$

其中, 每个 A_i 或者是不可约矩阵, 或者是 1×1 零矩阵, $i=1,\ \cdots,\ k$. 这种形式称为 A 的不可约正规形式, 注意, $\sigma(A)=\bigcup_{i=1}^{k}\sigma(A_i)$, 且 A 的不可约正规形式未必唯一.

9. 矩阵 $A=[a_{ij}]\in M_n(\mathbf{R})$称为本性非负矩阵, 是指其所有非对角元 $a_{ij}\,(i\neq j)$都是非负的. 证明, 如果 A 是本性非负矩阵, 则有某个 $\lambda>0$ 使得 $\lambda I+A\geqslant 0$. 试用这一论断以及(8.3.1)证明, 如果 $A\in M_n$ 是本性非负矩阵, 则 A 有实特征值 $r(A)$(常称其为 A 的优势特征值), 且有如下性质: 对 A 的每个特征值 λ_i 有 $r(A)\geqslant \mathrm{Re}\lambda_i$. 证明, $r(A)$不一定是 A 的具有最大模的特征值, 但当 $A\geqslant 0$ 时 $r(A)=\rho(A)$. **提示**: $\lambda I+A$ 的诸特征值为 $\lambda+\lambda_i$.

10. 定理(8.1.18)是说, 若 $A\in M_n$ 是非负矩阵, 则当 $B\in M_n$ 也是非负矩阵时有 $\rho(A+B)\geqslant\rho(A)$; 这是关于谱半径的一种单调性结果. 如果 $A\in M_n$ 是本性非负矩阵(见习题 9), 证明, 对所有对角矩阵 $D\in M_n(\mathbf{R})$, $A+D$ 是本性非负矩阵. 如果 A 是给定的本性非负矩阵, 而 D 允许在实对角矩阵类中变化, 则可得知, 优势特征值 $r(A+D)$是 D 的诸对角元的一个凸函数.

进一步阅读 参看以下文章: C. Johnson, R. Kellogg, and A. Stephens, "Complex

Eigenvalues of a Nonnegative Matrix with a Specified Graph II,"*Lin. Multilin. Alg.* 7 (1979), 129-143, 以及 C. Johnson, "Row Stochastic Matrices Similar to Doubly Stochastic Matrices," *Lin. Multilin. Alg.* 10 (1981), 113-130. 可以了解有关非负矩阵可能具有的特征值这一重要论题的一些结果. 这些文章给出一些参考文献. 其中包括 Dmitriev, Dynkin 和 Karpelevich 的经典工作，以及非负逆特征值问题. 后一个问题(刻划可以是非负矩阵的谱的复数集的特征)尚未解决. 关于习题 10 的其他信息可参看 J. Cohen, "Convexity of the Dominant Eigenvalue of an Essentially Nonnegative Matrix," *Proc. Amer. Math. Soc.* 81 (1981), 657-658. 也可参看(8.4)节 15 题. 506

8.4 不可约非负矩阵

如果能够对不含任何 0 元的矩阵证明一结论，那么这个结论常常可推广到不可约矩阵，这是一个颇具启发性的有用法则. 我们在第 6 章推广基本的 Geršgorin 定理时，已有这个法则的一个例证，现在要给出另一个例证. 其基本思想已在定理(6.2.24)中得到证明；这里重述其相关的部分.

8.4.1 引理 设 $A\in M_n$，且假定 $A\geqslant 0$. 那么，A 是不可约矩阵，当且仅当$(I+A)^{n-1}>0$.

练习 如果 $A\in M_n$，证明，A 是不可约的，当且仅当 A^T 是不可约的.

目前，我们还需要下述简单的结果.

8.4.2 引理 设 $A\in M_n$，且设 $\lambda_1,\cdots,\lambda_n$ 是 A 的特征值(计重特征值). 则 $\lambda_1+1,\cdots,\lambda_n+1$ 是 $I+A$ 的特征值且 $\rho(I+A)\leqslant 1+\rho(A)$. 如果 $A\geqslant 0$，则 $\rho(I+A)=1+\rho(A)$.

证明：设 $\lambda\in\sigma(A)$ 有重数 k，则 λ 是特征方程 $p_A(t)=\det(tI-A)=0$ 的 k 重根. 另外，因 $\det(tI-A)=\det[(t+1)I-(A+I)]$，所以 $\lambda+1$ 是 $p_{A+I}(s)=\det[sI-(A+I)]=0$ 的 k 重根. 于是 $\lambda_1+1,\cdots,\lambda_n+1$ 是 $A+I$ 的特征值. 因此，$\rho(I+A)=\max\limits_{1\leqslant i\leqslant n}|\lambda_i+1|\leqslant\max\limits_{1\leqslant i\leqslant n}|\lambda_i|+1=1+\rho(A)$. 但是，当 $A\geqslant 0$ 时，根据(8.3.1)，$1+\rho(A)$是 $I+A$ 的特征值，因此，在这种情形，$\rho(I+A)=1+\rho(A)$. □

练习 试说明，为了证明上述引理的前一部分，下述论证为什么是不完全的：如果 λ 是 A 的特征值，则存在某个向量 $x\neq 0$ 使得 $Ax=\lambda x$. 但是，$(A+I)x=(\lambda+1)x$，所以 $\lambda+1$ 是 $A+I$ 的特征值. 507

8.4.3 引理 如果 $A\in M_n$，$A\geqslant 0$，且对某个 $k\geqslant 1$ 有 $A^k>0$，则 $\rho(A)$是 A 的代数单重特征值.

证明：如果 $\lambda_1,\cdots,\lambda_n$ 是 A 的特征值，则 $\lambda_1^k,\cdots,\lambda_n^k$ 是 A^k 的特征值. 由定理(8.3.1)可知，$\rho(A)$是 A 的特征值，因此，假如 $\rho(A)$是 A 的重特征值，则 $\rho(A)^k=\rho(A^k)$是 A^k 的重特征值. 但这是不可能的，因为由定理(8.2.10)可知，$\rho(A^k)$是 A^k 的单重特征值.

现在将可看到，在多大程度上 Perron 定理可以推广到非负不可约矩阵. 把正矩阵的 Perron 结果推广到非负矩阵，是与 Frobenius 的名字分不开的.

8.4.4 定理 设 $A\in M_n$，且假定 A 是不可约非负矩阵. 则

(a) $\rho(A)>0$；

(b) $\rho(A)$是 A 的特征值；

(c) 存在正向量 x 使得 $Ax=\rho(A)x$；

(d) $\rho(A)$是 A 的代数(因而几何)单重特征值.

证明：推论(8.1.25)说明，即使在比不可约性更弱的条件下，(a)仍然成立. 根据定理(8.3.1)，论断(b)对所有非负矩阵成立，同时它保证存在适合 $Ax=\rho(A)x$ 的非负向量 $x\neq 0$. 另一方面，$(I+A)^{n-1}x=[1+\rho(A)]^{n-1}x$. 又根据引理(8.4.1)，矩阵$(I+A)^{n-1}$是正的，所以由(8.1.14)可知，向量$(I+A)^{n-1}x$一定是正的. 因此 $x=[1+\rho(A)]^{1-n}(I+A)^{n-1}x>0$. 为了证明(d)，利用引理(8.4.2)可证，如果 $\rho(A)$是 A 的重特征值，则 $1+\rho(A)=\rho(I+A)$是 $I+A$ 的重特征值. 但是 $I+A\geqslant 0$，且根据引理(8.4.1)，$(I+A)^{n-1}>0$，因此由引理(8.4.3)可知，$1+\rho(A)$一定是 $I+A$ 的单重特征值. □

该定理保证，不可约非负矩阵相应于 Perron 根的特征空间是一维的. 对于不可约非负矩

508 阵，其分量和为 1 的唯一正特征向量称为 Perron 向量.

因为不可约非负矩阵有正特征向量，所以(8.1)节末的诸结果可以应用于这类矩阵，其中特别重要的是谱半径的变分特征(8.1.32). 另外，A^T 不可约，当且仅当 A 不可约，所以不可约非负矩阵也有正左特征向量. 因此，定理(8.3.4)对不可约非负矩阵成立. 这个事实在定理(8.1.18)的下述推广中是关键的.

8.4.5 定理 设 $A, B\in M_n$. 假定 A 是不可约非负矩阵，且 $A\geqslant |B|$，则 $\rho(A)\geqslant\rho(B)$. 如果 $\rho(A)=\rho(B)$，又 $\lambda=e^{i\varphi}\rho(B)$是 B 的特征值，则存在 $\theta_1, \cdots, \theta_n\in\mathbf{R}$ 使得 $B=e^{i\varphi}DAD^{-1}$，其中 $D=\mathrm{diag}(e^{i\theta_1}, \cdots, e^{i\theta_n})$.

证明：由定理(8.1.18)已经知道，如果 $A\geqslant |B|$，则 $\rho(A)\geqslant\rho(B)$. 如果 $\rho(A)=\rho(B)$，则存在某个 $x\neq 0$，使得 $Bx=\lambda x$，且 $|\lambda|=\rho(B)=\rho(A)$. 因而

$$\rho(A)|x|=|\lambda x|=|Bx|\leqslant|B||x|\leqslant A|x|.$$

因为 A 是不可约矩阵，所以，根据定理(8.3.4)可知，$A|x|=\rho(A)|x|$. 因而 $|Bx|=|B||x|=A|x|$. 此外，定理(8.4.4)的(c)和(d)也蕴涵 $|x|>0$，又因为 $|B|\leqslant A$，所以由(8.1.15)和 $|B||x|=A|x|$ 的事实推出 $|B|=A$. 如果定义 $\theta_k\in\mathbf{R}$ 为 $e^{i\theta_k}=x_k/|x_k|$，$k=1, \cdots, n$，如果$\lambda\equiv e^{i\varphi}\rho(A)$，又如果令 $D\equiv\mathrm{diag}(e^{i\theta_1}, \cdots, e^{i\theta_n})$，则 $x=D|x|$ 且 $\lambda x=e^{i\varphi}\rho(A)D|x|=BD|x|=Bx$. 因此，$e^{-i\varphi}D^{-1}BD|x|=\rho(A)|x|=A|x|$. 这个恒等式连同 $|x|>0$ 和 $|e^{-i\varphi}D^{-1}BD|=A$ 的事实蕴涵$e^{-i\varphi}D^{-1}BD=A$. □

练习 对上述证明的最后一部分，给出其证明细节. **提示**：设 $C=e^{-i\varphi}D^{-1}BD$，且注意到

$$A|x|=C|x|=|C|x||\leqslant|C||x|=A|x|,$$

因而在三角不等式中等式成立. 辐角 $\arg(c_{ij}|x_j|)=$常数，$c_{ij}\geqslant 0$，且 $c_{ij}=a_{ij}$.

当 $A>0$ 时，由 Perron 定理可知，$\rho(A)$是 A 的有最大模的唯一特征值. 当 $A\geqslant 0$ 时，可能存在多个有极大模的特征值，不过，在这种情形，A 应该有特殊的形式，且这些特征值应该位于一个很规则的图形内.

8.4.6 推论 设 $A\in M_n$，假定 A 是不可约非负矩阵，又假定具有极大模特征值 $\rho(A)$的集合 $S=$

509 $\{\lambda_n=\rho(A), \lambda_{n-1}, \cdots, \lambda_{n-k+1}\}$恰好有 k 个互不相同的元. 则每个特征值$\lambda_i\in S$有代数重数 1，且

$$S = \{e^{2\pi i p/k}\rho(A): p = 0,1,\cdots,k-1\},$$

即这些极大模特征值正好是 k 个次单位根与 $\rho(A)$ 的乘积，另外，如果 λ 是 A 的任一特征值，则对所有 $p=0, 1, \cdots, k-1$，$e^{2\pi i p/k}\lambda$ 是特征值.

证明：对于 S 中的每个特征值，记 $\lambda_{n-p}=e^{i\varphi_p}\rho(A)$，$p=0, 1, \cdots, k-1$，也就是 $\varphi_p=\arg(\lambda_{n-p})$. 假定 $k>1$，如果有必要，可以重新标记诸特征值，如果还有必要，也可以重新定义辐角使得 $0=\varphi_0<\varphi_1<\varphi_2<\cdots<\varphi_{k-1}<2\pi$. 利用上述定理，且取 $B\equiv A$ 和 $\lambda\equiv\lambda_{n-p}$，得知，对 $p=0, 1, \cdots, k-1$ 有 $A=B=e^{i\varphi_p}D_pAD_p^{-1}$. 因 $D_pAD_p^{-1}$ 相似于 A，所以它与 A 有相同的特征值，因此这个恒等式说明，对任意 $p=0, 1, \cdots, k-1$，如果 A 的特征值集合在复平面内作角 φ_p 的旋转，则它仍变到自身；(只要证明 $\varphi_p=2\pi p/k$)这就是上述最后一个论断. 此外，已知 $\lambda_n=\rho(A)$ 是 A 的代数单重特征值(因为 A 是不可约的)，所以依次令 $p=1, 2, \cdots, k-1$，便得出所有 λ_{n-p} 也是代数单重特征值.

不过，还可以说得更详细一些. 因为对每个 $p=0, 1, \cdots, k-1$，$S=\{\lambda_n, \lambda_{n-1}, \cdots, \lambda_{n-k+1}\}=\{e^{i\varphi_p}\lambda_n, e^{i\varphi_p}\lambda_{n-1}, \cdots, e^{i\varphi_p}\lambda_{n-k+1}\}$，所以一定存在某个 $q=q(p)$ 使得 $\rho(A)=\lambda_n=e^{i\varphi_p}\lambda_q$；即对每个 p，都存在某个 $q=q(p)$ 使得 $\varphi_p=2\pi-\varphi_q$[即 $\varphi_p=-\varphi_q(\mathrm{mod}2\pi)$]，因而，$e^{-i\varphi_p}\rho(A)\in S$. 另外，如果重复定理(8.4.5)给出的关于 A 的表示式，并且对于任意选定的适合 $0\leqslant r, m\leqslant k-1$ 的 r, m，取 $B=e^{i\varphi_r}D_rAD_r^{-1}$ 和 $\lambda\equiv\lambda_{n-m}=e^{i\varphi_m}\rho(A)$，则得出

$$A=e^{i\varphi_r}D_r\{e^{i\varphi_m}D_mAD_m^{-1}\}D_r^{-1}=e^{i(\varphi_m+\varphi_r)}D_rD_mA(D_rD_m)^{-1}.$$

因此，根据上述相同的论证. 在复平面内，A 的特征值集合经 $\varphi_m+\varphi_r$ 的旋转变到自身. 特别是，$\lambda_ne^{i(\varphi_m+\varphi_r)}=e^{i(\varphi_m+\varphi_r)}\rho(A)$ 一定是 A 的(具有极大模的)特征值，所以对某个 $j=j(m, r)$，一定有 $\varphi_m+\varphi_r=\varphi_j(\mathrm{mod}2\pi)$.

考虑集合 $G\equiv\{\varphi_0=0, \varphi_1, \cdots, \varphi_{n-k-1}\}\subset[0, 2\pi)$. 上一段已提供了以下信息：(a)$0\in G$；(b)如果 $\varphi_i, \varphi_j\in G$，则 $\varphi_i+\varphi_j(\mathrm{mod}2\pi)\in G$；(c)如果 $\varphi_i\in G$，则 $-\varphi_i(\mathrm{mod}2\pi)\in G$. 另外显然有，(d)如果 $\varphi_i, \varphi_j\in G$，则 $\varphi_i+\varphi_j=\varphi_j+\varphi_i(\mathrm{mod}2\pi)$. 因此，$G$ 是恰好有 k 个元素的 Abel 群；群的运算是"关于模为 2π 的加法". 因为有限 Abel 群任一元素的阶一定整除群的阶，所以每个 $e^{i\varphi_m}$ 一定是 p 次单位根，其中 $p=p(m)$ 是 k 的因子. 另外，我们可以不用群的理论直接证明这一事实. 510

因为对某个 j 以及每个 r 和 m 有 $\varphi_m+\varphi_r=\varphi_j(\mathrm{mod}2\pi)$，用归纳法可得出，(令 $r=m=1$)对所有 $r=0, 1, 2, \cdots(\mathrm{mod}2\pi)$ 有 $r\varphi_1\in G$；即对所有 $r=0, 1, 2, \cdots$有 $e^{ir\varphi_1}\rho(A)\in S$. 但是，假如 $e^{i\varphi_1}$ 不是单位根，这便蕴涵 S 有无限多个不同的元素，而这是不可能的. 因此，对某个适合 $1<p\leqslant k$ 的 p，$e^{ip\varphi_1}=1$；可假定 p 是使这个式子成立的最小整数. 我们知道，$0<\varphi_1<\varphi_2<\cdots<\varphi_{n-k+1}<2\pi$，且对某个固定指数 m 考虑 φ_m，恰好可用 $p+1$ 个点 $0, \varphi_1, 2\varphi_1, \cdots, (p-1)\varphi_1, p\varphi_1=2\pi$ 将区间$[0, 2\pi)$分成 p 个半开子区间(在右边开)；点 φ_m 一定落在这些子区间的一个中. 于是，存在某个适合 $0\leqslant q\leqslant p-1$ 的 q 使得 $q\varphi_1\leqslant\varphi_m<(q+1)\varphi_1$；即 $0\leqslant\varphi_m-q\varphi_1<\varphi_1$. 因此，对某个 $j=j(m)$，一定有 $\varphi_m-q\varphi_1=\varphi_j$，这是因为已经证明，如果 $e^{i\varphi_1}\rho(A)$是特征值，则 $e^{-i\varphi_1}\rho(A)$，$e^{-iq\varphi_1}\rho(A)$和 $e^{-iq\varphi_1+i\varphi_m}\rho(A)$也是特征值. 另一方面，$0\leqslant\varphi_m-q\varphi_1=\varphi_j<\varphi_1$，且选定 φ_1 是最小非零辐角，因而一定有 $\varphi_m-q\varphi_1=0$. 这就说明每个辐角 φ_j 是 φ_1 的某个倍数，所以必定是 $p=k$；即 $\varphi_1=2\pi/k$，因为假如 $p<k$，则在集合$\{e^{i\varphi_1}\rho(A), e^{2i\varphi_1}\rho(A), e^{3i\varphi_1}\rho(A), \cdots\}$中的不同的元素会少于 k

个，然而这个集合一定等于整个 S. 最后，因为每个 φ_m 是 $\varphi_1=2\pi/k$ 的某个倍数，又因为存在 k 个不同的 φ_i 项和 k 个 φ_1 的不同倍数，所以对所有 $m=0, 1, 2, \cdots, k-1$ 一定有 $\varphi_m=m\varphi_1$.

整个证明是在 $k>1$ 的假定下进行的，但是，如果 $k=1$，所作的论断是显然的. □

8.4.7 附注 如果 $A\geqslant 0$ 是不可约的，且有 $k>1$ 个极大模特征值，则 A 的每个非零特征值位于中心在 $\mathbf{C}$ 中的 0 的一个圆上，这个圆恰好经过 A 的 k 个特征值，且使它们在整个圆上都保持相同的间隔. 特别是，k 必须是 A 的非零特征值的个数的因子. 因此，如果 A 是 $n\times n$ 非奇异不可约非负矩阵，且 n 是素数，则一定存在一个或 n 个极大模特征值；不会有其他可能性.

8.4.8 推论 假定 $A\in M_n$ 是不可约非负矩阵，且记 $A^m=[a_{ij}^{(m)}]$，$m=1, 2, \cdots$. 如果正好存在 A 的 $k>1$ 个具有极大模的特征值，则当 m 不是 k 的整数倍时，$a_{ii}^{(m)}\equiv 0$ 对所有 $i=1, 2, \cdots, n$ 成立. 特别是所有 $a_{ii}=0$.

511

证明：利用推论(8.4.6)选取 A 的一个极大模特征值 $\lambda=e^{i\varphi}\rho(A)$，其中 $\varphi=2\pi/k$. 于是，只要 m 不是 k 的整数倍，$e^{im\varphi}$ 就不是正实数. 利用定理(8.4.5)，且取 $B=A$ 和 $\lambda=e^{i\varphi}\rho(A)$，得出 $A=e^{i\varphi}DAD^{-1}$，所以对于所有 $i=1, \cdots, n$ 和 $m=1, 2, 3, \cdots$，有 $A^m=e^{im\varphi}DA^mD^{-1}$ 和 $a_{ii}^{(m)}=e^{im\varphi}a_{ii}^{(m)}$. 如果 $e^{im\varphi}$ 不是正实数，又如果 $a_{ii}^{(m)}>0$，则这个等式不可能成立，因此，当 m 不是 k 的倍数时，对所有 $i=1, \cdots, n$，一定有 $a_{ii}^{(m)}\equiv 0$. □

练习 假定 $A\in M_n$ 是不可约非负矩阵. 证明，为了保证 $\rho(A)$ 是 A 的唯一极大模特值，其充分条件是某个 $a_{ii}\neq 0$. 但是，考察矩阵

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

并说明这个条件不是必要的，你能找一个 2×2 矩阵的反例吗？

8.4.9 附注 比推论(8.4.8)更强的结论仍成立：如果 $A\geqslant 0$ 是不可约矩阵，且有 $k>1$ 个具有极大模的特征值，则存在置换矩阵 P 使得

$$PAP^T=\begin{bmatrix} 0 & A_{12} & & 0 \\ \vdots & 0 & \ddots & \\ 0 & & \ddots & A_{k-1,k} \\ A_{k,1} & 0 & \cdots & 0 \end{bmatrix}$$

其中，k 个主对角零子块是方阵而所列出的诸子块 A_{ij} 不一定是零矩阵. 特别地，所有主对角元 a_{ii} 必须是零[Var, p. 28].

为了得到在推论(8.4.6)中所描述的极大模特征值的规则图形，尽管不可约性假设是本质的，但是在一般情形下不仍然可以得到某些信息.

8.4.10 推论 如果 $A\in M_n$，$A\geqslant 0$，$\rho(A)>0$，如果 λ 是 A 的适合 $|\lambda|=\rho(A)$ 的特征值，则 $\lambda/\rho(A)=e^{i\varphi}$ 是单位根，对某个适合 $1\leqslant k\leqslant n$ 的 k 有 $e^{ik\varphi}=1$，且对所有 $p=0, 1, 2, \cdots, k-1$，$e^{ip\varphi}\rho(A)$ 都是 A 的特征值.

512

证明：如果 A 是不可约的，则论断可由推论(8.4.6)推出. 如果 A 可约，则 A 可经置换相似变成分块上三角形式

$$\begin{bmatrix} A_1 & & & * \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_r \end{bmatrix}$$

其中A_j是不可约方阵或零方阵. A的特征值是各对角子块矩阵A_1，…，A_r的特征值的并，且每个A_j的极大模特征值集合的结构由推论(8.4.8)给出. □

练习 设

$$A_1=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad 和 \quad A_2=\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix},$$

通过考察

$$A=\begin{bmatrix} A_1 & * \\ 0 & A_2 \end{bmatrix}\in M_5$$

来说明，对于一般的非负矩阵A，A的极大模特征值可能比$\rho(A)$以及$\rho(A)$经过一个单位根的各次幂的旋转所得到的集合要多.

习题

1. 用例子说明，未列入到定理(8.4.4)中的Perron定理(8.2.11)中的各项命题对不可约非负矩阵一般不成立.

2. 用例子说明，$\rho(I+A)=1+\rho(A)$对所有$A\in M_n$不成立. 试给出使这恒等式成立的关于A的必要充分条件.

3. 不可约性是非负矩阵有正特征向量的充分条件而不是必要条件. 试考察$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$和$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$，说明可约非负矩阵可能有或可能没有正特征向量.

4. 如果$A\geqslant 0$是不可约的，证明，当$m\to\infty$时，矩阵$[\rho(A)^{-1}A]^m$的各元一致有界.

513

5. 我们已经证明，不可约矩阵有正Perron向量. 假定$A\geqslant 0$，$\rho(A)>0$，$x\geqslant 0$，$x\neq 0$且$Ax=\rho(A)x$；证明，如果x不是正的，则A是可约的. 如果x是正的，A就一定不可约吗？

6. 假定$A\geqslant 0$是不可约的，且$B\geqslant 0$与A可交换. 如果x是A的Perron向量，证明$Bx=\rho(B)x$. **提示：**定理(8.4.4d).

7. 说明多项式$x^k-1=0$的友矩阵是$k\times k$非负矩阵有k个极大模特征值的例子. 验证这些特征值在复平面中的位置.

8. 设k_1，k_2，k_p是给定的正整数. 说明如何构造一个非负矩阵使得它的极大模特征值正好是k_1个k_1次单位根，k_2个k_2次单位根，…，以及k_p个k_p次单位根.

9. 如果不可约非负矩阵A有$k\geqslant 1$个极大模特征值，说明为什么称A为指标k的循环.

10. 如果$A\geqslant 0$是不可约矩阵，并且还是指标$k\geqslant 1$的循环，证明特征多项式$p_A(t)=t^r(t^k-\rho(A)^k)(t^k-\mu_2^k)\cdots(t^k-\mu_m^k)$，其中，$r$，$m\geqslant 0$为某个整数，$\mu_i$是适合$|\mu_i|<\rho(A)$的某些复数，$i=2$，…，$m$. 对$p_A(t)$中零和非零系数的型式作出说明，并且基于特征多项式的形式给出A只有一个极大模特征值的准则. **提示：**在推论(8.4.6)的证明中已知，如果$\varphi=2\pi/k$，

且 λ 是 A 的特征值，则 $e^{ir\varphi}\lambda$ 亦是 A 的特征值，$r=0, 1, 2, \cdots$.

11. 设 $n>1$ 是素数. 证明，如果 $A\in M_n$ 是非负的，不可约的和非奇异的，则或者 $\rho(A)$ 是 A 的有极大模的唯一特征值，或者 A 的所有特征值有极大模.

12. 考察 $A=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$，说明一般不可能改进推论(8.4.8)而断言，$A$ 的所有幂的主对角元一定都是零.

13. 如果 $A\geqslant 0$，证明，A 的不可约性只与其零元的位置有系，而与非零元的大小无关.

14. 如果 A，$B\in M_n$，则 AB 与 BA 有相同的特征值. 考察 $\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$ 和 $\begin{bmatrix}0 & 0\\ 1 & 1\end{bmatrix}$，说明，即使 A 和 B 是非负矩阵，有可能 AB 不可约而 BA 可约. 这种例子说明不可约矩阵可能相似(甚至酉等价)于可约矩阵；解释这是为什么. 同时它说明，仅仅涉及矩阵的特征值的任何条件不能最终验证其不可约性.

15. 设 $A\in M_n$ 是给定的不可约非负矩阵. 证明，只要 $B\in M_n$ 非负，$A+B$ 就不可约，并且只要 $B\geqslant 0$ 和 $B\neq 0$，就有 $\rho(A+B)>\rho(A)$. 这是(8.1.18)到严格单调性的改进，不过需附加不可约性假定. **提示**：(8.1.18)说明 $\rho(A+B)\geqslant\rho(A)$. 如果等式成立，利用(8.4.5)证明 $B=0$.

16. 证明可以用下述方式强化(8.4.1). 设 $A\in M_n$ 是非负的，且设 A 的极小多项式有次数 m. 证明，A 不可约，当且仅当 $(I+A)^{m-1}>0$. **提示**：考虑 $I+A+A^2+\cdots+A^{m-1}+A^m+\cdots+A^{n-1}$，然后用 I，A，$\cdots$，A^{m-1} 表示 A^m 和更高次幂.

17. 设 $A\in M_n$ 是给定的非负矩阵，考虑在最小二乘意义下求 A 的秩 1 最佳逼近问题；即求秩 1 $X\in M_n$ 使 $\| A-X \|_2=\min\{\| A-Y \|_2 : Y\in M_n$ 有秩 1$\}$. 假定 AA^T 的 Perron 根是单根，如果 AA^T 或 A^TA 不可约，情况就是这样. 为什么？证明这样的最佳 X 是非负的，唯一的且由 $X=\sqrt{r}vw^T$ 给出，其中 $r=\rho(AA^T)$ 是 A 的 Perron 根，而 v，$w\in\mathbf{R}^n$ 是非负单位向量，它们分别是 AA^T 和 A^TA 相应于特征值 r 的单位特征向量. **提示**：利用(7.4.1)中给出的秩 1 最佳逼近的特征. 注意 AA^T 和 A^TA 都是实对称半正定矩阵，所以 r，v 和 w 的计算原则上是不太困难的.

18. 试用习题 17 求每个矩阵

$$A=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\quad \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix},\quad \begin{bmatrix}0 & 0\\ 1 & 1\end{bmatrix}$$

的秩 1 最佳二乘逼近. 证明，$A=I\in M_n$ 的秩 1 最佳二乘逼近不唯一；对于任何单位向量 $v\in\mathbf{C}^n$，$X=vv^*$ 是 I 的秩 1 最佳二乘逼近.

8.5 素矩阵

实际上，在 Perron 定理中可能用得最多的结果是定理(8.2.8)中的极限命题. 定理(8.4.4)的证明说明，没有把引理(8.2.7)应用于不可约矩阵的仅有假设条件是谱半径为唯一的极大模特征值. 因为 $A=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$ 是有两个极大模特征值的非负不可约矩阵的例子(并且对于它，$\lim_{m\to\infty}A^m$ 不存在)，所以对不可约矩阵类作进一步的限制是必要的；而最经济的办法是缺什

么就假定什么.

8.5.0 定义 非负矩阵 $A\in M_n$ 称为素矩阵(或本原矩阵)，是指它是不可约的且只有一个极大模特征值.

素性的概念属于 Frobenius (1912). 现在，采用与定理(8.2.8)相同的证明可直接由引理(8.2.7)得到下述极限结果.

8.5.1 定理 如果 $A\in M_n$ 是非负素矩阵，则

$$\lim_{m\to\infty}[\rho(A)^{-1}A]^m = L > 0,$$

其中，$L=xy^T$，$Ax=\rho(A)x$，$A^Ty=\rho(A)y$，$x>0$，$y>0$ 且 $x^Ty=1$. 另外，如果 λ_{n-1} 是 A 的特征值，且对每个特征值 $\lambda\neq\rho(A)$ 有 $|\lambda|\leqslant|\lambda_{n-1}|$，又如果 $|\lambda_{n-1}|/\rho(A)<r<1$，则存在常数 $C=C(r, A)$，使得 $\|[\rho(A)^{-1}A]^m-L\|_\infty\leqslant Cr^m$ 对所有 $m=1, 2, \cdots$成立. □

我们现在已经把 Perron 定理的所有结果从正矩阵类推广到非负素矩阵类. 但是，实际上仍然需要解决验证一个已知非负矩阵的素性问题，从理论上讲，可能希望不直接计算特征值就能做到这一点. 下面的素性特征诱导出几个有用的准则，而其特征在计算上并不是有效的检验法.

8.5.2 定理 如果 $A\in M_n$ 是非负的，则 A 是素矩阵，当且仅当 $A^m>0$ 对某个 $m\geqslant1$ 成立.

证明： 如果 $A\geqslant0$，且 $A^m>0$，则从 A 的有向图 $\Gamma(A)$的每个结点 P_i 到另一个结点 P_j 一定存在一条长恰好为 m 的有向道路(推论(6.2.18)). 因为这是比不可约性更强的性质，所以 A 一定不可约. 像在(8.4.3)中那样应用 Perron 定理(8.2.11 d，e)，则推出 A 一定是素的. 反之，如果 A 是素矩阵，则根据定理(8.5.1)有 $\lim_{m\to\infty}[\rho(A)^{-1}A]^m=L>0$，且对某个 $m\geqslant1$ 必定有 $[\rho(A)^{-1}A]^m>0$. □ 516

现在联想起不可约性的图解准则，上述特征连同已有的关于非负不可约矩阵的极大模特征值的非常强的信息给我们提供了素性的图解准则. 我们知道，正整数序列 k_1，k_2，…的最大公因子(g. c. d.)是使 k 为所有 k_1，k_2，…的因子的最大整数 $k\geqslant1$.

8.5.3 定理 设 $A\in M_n$ 是不可约非负矩阵，设$\{P_i\}$表示有向图 $\Gamma(A)$的结点集. 用 $L_i=\{k_1^{(i)}, k_2^{(i)}, \cdots\}$表示 $\Gamma(A)$中起点和终点都在结点 P_i($i=1, 2, \cdots, n$)的所有有向道路的长度集合. 用 g_i 表示 L_i 中的所有长度的最大公因子. 则 A 是素矩阵，当且仅当所有 $g_i=1$，$i=1, 2, \cdots, n$.

证明： 我们知道，因为 A 不可约，所以其长度的集合 L_i 非空；对每个 i 和任意 $j\neq i$，在 $\Gamma(A)$中有一条连结 P_i 到 P_j 的道路，同时在 $\Gamma(A)$中也有一条连结 P_j 到 P_i 的道路. 如果 A 是素矩阵，则根据定理(8.5.2)，存在某个 $m\geqslant1$ 使得 $A^m>0$，因而对所有 $k\geqslant m$ 有 $A^k>0$. 另一方面，对所有 $i=1, \cdots, n$，有 m，$m+1$，$m+2$，$\cdots\in L_i$，因而 $g_i=1$ 对所有 $i=1, \cdots, n$ 成立.

现在假设 $A=[a_{ij}]$不是素矩阵. 如果 A 恰好有 $k>1$ 个极大模特征值，则由推论(8.4.8)可知，对所有 $i=1, \cdots, n$ 以及所有不是 k 的倍数的 m 有 $a_{ii}^{(m)}=0$. 因此 $L_i\subset\{k, 2k, 3k, \cdots\}$，因而 $g_i\geqslant k>1$ 对所有 $i=1, \cdots, n$ 成立. □

8.5.4 附注 比定理(8.5.3)中的论断稍微进一步的结论也成立；事实上，总有 $g_1=g_2=\cdots=g_n$，而 g_i 个项的公共值恰好是 A 的极大模特征值的个数．这就是 Romanovsky 定理．

下面的结果在许多场合是有用的；特别是它证明了具有正主对角元的不可约非负矩阵一定是素矩阵．

8.5.5 引理 如果 $A\in M_n$ 是不可约非负矩阵，又如果 A 的所有主对角元都是正的，则 $A^{n-1}>0$．

证明：如果 $\alpha=\min\{a_{11}, a_{22}, \cdots, a_{nn}\}$，又如果定义

517

$$B = A - \operatorname{diag}(a_{11}, a_{22}, \cdots, a_{nn}),$$

则 B 是不可约非负矩阵(因为 A 是不可约的)，且 $A\geqslant\alpha I+B=\alpha[I+(1/\alpha)B]$，因而由引理(8.4.1)可知，$A^{n-1}\geqslant\alpha^{n-1}[I+(1/\alpha)B]^{n-1}>0$． □

练习 试说明，当对具有正对角元的非负方阵取升幂时，变成正元的任何元在所有各次幂中仍然是正的．

虽然不可约矩阵的幂可能是可约的，但是素矩阵的所有幂都是素矩阵．

8.5.6 引理 设 $A\in M_n$ 是非负素矩阵．则对所有 $k=1, 2, \cdots$，A^k 是不可约非负素矩阵．

证明：因为 A 的所有充分大的幂是正的，所以对任意 k，A^k 也是正的．假如对某个 k，A^k 是可约的，则 A^k 的所有幂也是可约的，因而不可能是正的．因为这与 A 的所有充分大的幂是正的事实相矛盾，所以对 A 的任何次幂都不可能是可约的． □

从计算上考虑，定理(8.5.2)中的特征本身不是验证素性的有效方法．因为要计算的幂没有给出其上界．如果求出 m 使 $A^m>0$，则 A 是素矩阵；但是，如果我们还没有得到一个正幂，那么什么时候终止这种计算呢？下述定理给出的有限界回答了这个问题．

8.5.7 定理 设 $A\in M_n$ 是非负矩阵．如果 A 是素矩阵，则对某个正整数 $k\leqslant(n-1)n^n$ 有 $A^k>0$．

证明：因为 A 是不可约的，则在 $\Gamma(A)$ 中存在从结点 P_1 回到结点 P_1 的有向道路；这样的最短道路有长 $k_1\leqslant n$．因此矩阵 A^{k_1} 在它的 1，1 位置有正元，并且 A^{k_1} 的任意幂也将有正 1，1 元．因为 A 是素的，由引理(8.5.6)可知 A^{k_1} 一定不可约，因而在 $\Gamma(A^{k_1})$ 中一定存在从结点 P_2 回到结点 P_2 的有向道路；这样的最短道路有长 $k_2\leqslant n$．因此矩阵 $(A^{k_1})^{k_2}=A^{k_1k_2}$ 有正 1，1 元和正 2，2 元．这个过程可沿主对角线继续进行下去直到得到矩阵 $A^{k_1k_2\cdots k_n}$(其中每个 $k_i\leqslant n$)为止，

518

这个矩阵是不可约的且有正对角元，因而由引理(8.5.5)可知$[A^{k_1k_2\cdots k_n}]^{n-1}>0$．因为

$$k_1k_2\cdots k_n(n-1)\leqslant n\cdot n\cdots n(n-1)=n^n(n-1),$$

所以证毕． □

如果 A 是给定的素矩阵，使 $A^k>0$ 的最小 k 称为 A 的本原指标，通常记作 $\gamma(A)$．我们已经看到，如果 A 有正对角元，则 $\gamma(A)\leqslant n-1$，而在一般情形下有 $\gamma(A)\leqslant n^n(n-1)$，后一个界可以得到显著改进．

8.5.8 定理 设 $A\in M_n$ 是非负素矩阵，且假定 $\Gamma(A)$ 中的最短简单有向回路有长 s．则 $A^{n+s(n-2)}>0$，即 $\gamma(A)\leqslant n+s(n-2)$．

证明：因为 A 是不可约的，所以 $\Gamma(A)$ 的每个结点落在一条回路上，且从任一结点回到自身的最短回路将是长度至多为 n 的简单回路．经过一个置换，可以假定，在这条最短回路中的结点是 P_1，P_2，…，P_s．注意到 $n+s(n-2)=n-s+s(n-1)$ 且考虑 $A^{n-s+s(n-1)}=A^{n-s}(A^s)^{n-1}$．把 A^{n-s} 写成分块形式

$$A^{n-s}=\begin{bmatrix}X_{11} & X_{12}\\ X_{21} & X_{22}\end{bmatrix},$$

其中 $X_{11}\in M_s$ 和 $X_{22}\in M_{n-s}$．因为结点 P_1，…，P_s 组成 $\Gamma(A)$ 中的一条回路，所以 X_{11} 在每一行至少有一个非零元，因而从图 $\Gamma(A^{n-s})$ 中的每个结点 P_i 到 $\Gamma(A^{n-s})$ 中的某个结点 P_j（或许 $P_i=P_j$）存在某条弧；如果 $1\leqslant i$，$j\leqslant s$，这是成立的．因为从不在上述回路中的每个结点 P_{s+1}，…，P_n，到该回路中的某个结点，在 $\Gamma(A)$ 中一定有一条长不超过 $n-s$（它们是不在该回路中的结点数）的有向道路，所以在 X_{21} 的每一行中至少有一个非零元．当绕着回路走了足够多的附加步数时，从不在该回路中的每个结点到该回路的某个结点，显然在 $\Gamma(A)$ 中恰好有一条长恰好为 $n-s$ 的有向道路．

现在把 $(A^s)^{n-1}$ 写成如下的分块形式

$$(A^s)^{n-1}=\begin{bmatrix}Y_{11} & Y_{12}\\ Y_{21} & Y_{22}\end{bmatrix},$$

其中，$Y_{11}\in M_s$，而 $Y_{22}\in M_{n-s}$．因为 P_1，…，P_s 组成 $\Gamma(A)$ 中的一条回路，所以在 $\Gamma(A^s)$ 中的每个结点 P_1，…，P_s 有一个圈．因为 A 是素矩阵，所以 A^s 也是素矩阵，因而是不可约的．从 $\Gamma(A^s)$ 的每个结点 P_1，…，P_s 到任意其他结点，在 $\Gamma(A^s)$ 中有一条长至多为 $n-1$ 的道路．首先绕着起点的圈走足够多次，便可以构造这样一条长恰好为 $n-1$ 的道路．这说明 $Y_{11}>0$ 和 $Y_{12}>0$． 519

为了完成证明，算出

$$A^{n-s}(A^s)^{n-1}=\begin{bmatrix}X_{11} & X_{12}\\ X_{21} & X_{22}\end{bmatrix}\begin{bmatrix}Y_{11} & Y_{12}\\ Y_{21} & Y_{22}\end{bmatrix}\geqslant\begin{bmatrix}X_{11}Y_{11} & X_{11}Y_{12}\\ X_{21}Y_{11} & X_{21}Y_{12}\end{bmatrix}.$$

因为在后一个表示式中的 X 子块的每一行至少含有一个非零元，又因为在后一个表示式中的每个 Y 子块都是正的，所以整个分块矩阵是正的，且 $A^{n-s}(A^s)^{n-1}>0$． □

(8.5.8)的一个推论是 H・Wielandt 的著名结果，它给出了一般的素矩阵的本原指标的准确上界．

8.5.9 推论 如果 $A\in M_n$ 是非负矩阵，则 A 是素矩阵，当且仅当 $A^{n^2-2n+2}>0$．

证明：如果 A 的任意幂是正的，则 A 是素矩阵，所以，我们只要考虑逆命题．如果 $n=1$，结论是显然的，因此假定 $n>1$．如果 A 是素矩阵，则 A 是不可约的，且在 $\Gamma(A)$ 中有若干条回路．如果 $\Gamma(A)$ 中最短的回路有长 n，则 $\Gamma(A)$ 中的每条回路的长是 n 的倍数，因而由定理(8.5.3)可知，A 不能是素矩阵．于是 $\Gamma(A)$ 中最短回路的长是 $n-1$ 或者更小，因此根据定理(8.5.8)有

$$\gamma(A)\leqslant n+s(n-2)\leqslant n+(n-1)(n-2)=n^2-2n+2.$$

□

Wielandt 给出的一个例子(本节末习题 4)说明，界 $\gamma(A)\leqslant n^2-2n+2$ 是所有对角元全为 0 的矩阵的最佳界. 我们知道，如果所有主对角元是正的，则 A 是素矩阵，当且仅当 $A^{n-1}>0$. 如果某些主对角元为正，可能不是全部主对角元为正，则 Holladay 和 Varga 的下述结果利用定理(8.5.8)中所采用的证明思想给出了关于本原指标的一个界.

8.5.10 定理 设 $A\in M_n$ 是不可约非负矩阵，且假定 A 有 d 个正主对角元，$1\leqslant d\leqslant n$. 则 $A^{2n-d-1}>0$；即 $\gamma(A)\leqslant 2n-d-1$.

证明：在所述假设下，A 一定是素矩阵，且在 $\Gamma(A)$ 中极小长回路有长 1，实际上有 d 条这样的回路. 通过置换，可以假定 $P_1,\cdots,P_d$ 是 $\Gamma(A)$ 中的具有圈的结点. 考虑 $A^{2n-d-1}=$
520
$A^{n-d}(A^1)^{n-1}$，且记

$$A^{n-d}=\begin{bmatrix}X_{11} & X_{12}\\ X_{21} & X_{22}\end{bmatrix},\quad A^{n-1}=\begin{bmatrix}Y_{11} & Y_{12}\\ Y_{21} & Y_{22}\end{bmatrix},$$

其中，X_{11}，$Y_{11}\in M_d$，而 X_{22}，$Y_{22}\in M_{n-d}$. 采用在定理(8.5.8)证明中的相同证法来处理 A^{n-d} 和$(A^1)^{n-1}$的对应位置的子块便可证明，子块 X_{11} 和 X_{21} 的每一行至少含有一个非零元，且子块 Y_{11} 和 Y_{12} 是正的. 由定理(8.5.8)中所运用的相同推理可知，乘积 $A^{n-d}A^{n-1}$ 是正矩阵. □

练习 证明矩阵 $A=\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}$是素矩阵. 它的特征值是什么？计算出(8.5.9)和(8.5.10)给出的关于 $\gamma(A)$的界. $\gamma(A)$的精确值是什么？

作为最后一个附注，我们指出，如果想证明一个给定的非负矩阵是素的，则可以验证该矩阵是不可约的且满足 Wielandt 条件(8.5.9). 在实际中出现的矩阵常常有特殊的结构，使我们容易看出相应的有向图是不是强连接的. 另外，如果任一主对角元是正的，则该矩阵一定是素的. 但是，如果矩阵很大，且没有特殊的结构或者其元素没有什么对称性，或者如果所有主对角元是零，则可能需要利用定理(8.4.1)或推论(8.5.9)来验证不可约性或素性. 在这两种情形下，如果将所述矩阵反复平方若干次直到所得到的幂超过临界值(分别为 $n-1$ 或 n^2-2n+2)为止，则矩阵乘法所需要的次数将大大缩小. 例如，如果 $n=10$，则计算$(I+A)^2$，$(I+A)^4$，$(I+A)^8$ 和$(I+A)^{16}$足以验证不可约性；这是 4 次矩阵乘法而不是直接应用引理(8.4.1)所需要的 8 次乘法. 类似的，如果 A 是非负的，则计算 A^2，A^4，A^8，A^{16}，A^{32}，A^{64}和 A^{128}足以验证素性；这是 7 次矩阵乘法而不是 81 次乘法. 注意在这些讨论中我们未指明利用了习题 3.

习题

1. 写出定理(8.5.1)的证明.

2. 若 $A\in M_n$ 是非负素矩阵，证明对所有 i，$j=1,\cdots,n$ 有 $\lim\limits_{m\to\infty}[a_{ij}^{(m)}]^{1/m}=\rho(A)$. 试把这
521
个结果与推论(5.6.14)作一比较. 可以省略掉有关素性假设的那一部分吗？

3. 证明，如果 $A\geqslant 0$ 和 $A^k>0$，则 $A^m>0$ 对所有 $m\geqslant k$ 成立. 如果 A 是素矩阵，则对任意正整数 k，A^k 是素矩阵. 但是，如果 A 和 B 都是素矩阵，则可能 AB 不是素矩阵. **提示：** 考虑$\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}$和$\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}$.

4. 试用 $\Gamma(A)$证明，对 $n\geqslant 3$，Wielandt 矩阵

$$A=\begin{bmatrix} 0 & 1 & & & & \\ 0 & & 1 & & 0 & \\ \vdots & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ 0 & & 0 & & & 1 \\ 1 & & 1 & 0 & \cdots & 0 \end{bmatrix}\in M_n$$

是不可约素矩阵. 然后证明 A^{n^2-2n+1} 的(1, 1)元是零而 $A^{n^2-2n+2}>0$. **提示**：把 A 看成作用在标准基$\{e_1, \cdots, e_n\}$上的线性变换. 那么 A：$e_1\to$? A^{n-1}：$e_1\to$? $A^{(n-1)(n-1)}$：$e_1\to$?

5. 设 $A\in M_n$ 是不可约非负矩阵. 证明，如果至少有一个主对角元是正的，则 A 是素矩阵. 证明这个充分条件对 $n=2$ 是必要条件而对 $n\geqslant 3$ 则不是.

6. 设 $A=[a_{ij}]\in M_n$ 是非负的，且假定 $a_{kk}>0$ 对某个 $k=1, 2, \cdots, n$ 成立. 证明 A 的每个幂的 k，k 元也是正的. 如果 $a_{kk}=0$ 而 A^2 的 k，k 元是正的，A^3 的 k，k 元是正的吗?

7. 详细验证本节末提出的简便算法.

8. 如果 A 是任意幂等矩阵，则 $A=\lim\limits_{m\to\infty}A^m$. 证明，如果 A 是非负的，不可约的和幂等的，则 A 是秩 1 的正矩阵.

9. 给出例子说明，即使 $A\geqslant 0$ 不是素矩阵，$\lim\limits_{m\to\infty}[\rho(A)^{-1}A]^m$ 仍可能存在. 实际上，A 可能是可约的，还可能有极大模重特征值.

10. 证明定理(8.5.1)的下述部分逆命题：如果 $A\in M_n$ 是非负矩阵和不可约矩阵，又如果 $\lim\limits_{m\to\infty}[\rho(A)^{-1}A]^m$ 存在，则 A 是素矩阵. **提示**：如果 $|\mu|=\rho(A)$，$\mu\neq\rho(A)$，且 $Az=\mu z$，$z\neq 0$，则$[\rho(A)^{-1}A]^m z\to$?

11. 证明. $A=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$是不可约的，但 A^2 是可约的. 这与(8.5.6)矛盾吗?

522

12. 给出一个不可约非负矩阵 $A\in M_n$，使得$\lim\limits_{m\to\infty}[\rho(A)^{-1}A]^m$ 不存在.

13. 如果 $\varepsilon>0$，又如果 $A\in M_n$ 是非负不可约矩阵，证明 $A+\varepsilon I$ 是素矩阵.

14. 非负矩阵 $A=[a_{ij}]$称为组合对称的，是指 $a_{ij}>0$ 当且仅当 $a_{ji}>0$ 对所有 i，$j=1, \cdots, n$ 成立. 证明，如果 A 是组合对称的素矩阵，则 $A^{2n-2}>0$. **提示**：考虑 A^2 且利用(8.5.6)和(8.5.10). 你能给出关于 $\Gamma(A)$的回路结构的更多的信息来改进 $\gamma(A)$的界吗? **提示**：利用(8.5.8).

15. 证明，如果 $A\in M_n$ 是非负的，不可约的和非奇异的，且 n 是素数，则或者(a)A 是素矩阵，或者(b)A 的所有特征值有极大模且 A 相似于 $x^n-\rho(A)^n=0$ 的友矩阵.

16. 一种计算非负矩阵 $A\in M_n$ 的 Perron 向量和谱半径的方法是幂法：

$$x^{(0)}\text{ 是任意正向量，}\quad \sum_{i=1}^{n}x_i^{(0)}=1;$$

$$y^{(m+1)}=Ax^{(m)}\quad\text{对所有 } m=0,1,2,\cdots\text{ 成立;}$$

$$x^{(m+1)} = \frac{y^{(m+1)}}{\sum_{i=1}^{n} y_i^{(m+1)}}$$ 对所有 $m = 0,1,2,\cdots$ 成立.

如果 A 是素矩阵，证明，向量 $x^{(m)}$ 的序列收敛于 A 的(右)Perron 向量，而数 $\sum_{i=1}^{n} y_i^{(m+1)}$ 的序列收敛于 A 的 Perron 根. 其收敛速度如何? 关于素性的假设是必要的吗?

17. 如果 $A \in M_n$ 是非负的，证明 A 的素性只与零元的位置有关而与非零元的大小无关.

18. 如果 $A \in M_n$ 是非负的、不可约的和对称的，证明 A 是素矩阵当且仅当 $A + \rho(A)I$ 是非奇异矩阵. 特别是，如果 A 是半正定矩阵，这个条件被满足. 以 0 和 1 为元素的对称非负矩阵常常作为无向图的邻接矩阵而出现.

523

19. 如果 $A \in M_n$ 是素矩阵且 $k \geq \gamma(A)$，证明 $A^k > 0$.

20. 给出定理(8.5.10)的证明细节.

21. 计算下列各矩阵的特征值和特征向量，并且按本章的基本概念(非负矩阵、不可约矩阵、素矩阵、正矩阵等)进行归类：$\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$, $\begin{bmatrix}0 & 1\\1 & 1\end{bmatrix}$, $\begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}$, $\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$, $\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$, $\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}$, $\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$. 这些矩阵为可能出现的各种可能性提供了好的实例.

22. 在定理(8.5.8)的证明中，证明 X_{11} 和 X_{12} 的每一列至少含有一个非零元. 证明 $Y_{21} > 0$.

进一步阅读 关于(8.5.4)中提到的 Romanovsky 定理的证明可参看 V. Romanovsky, "Recherches sur les Chaines de Markoff," *Acta Math*. 66(1936), 147-251.

8.6 一般极限定理

即使非负矩阵 A 是不可约的，A 的标准幂却不一定有极限，以矩阵

$$A = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$$

为例就不难说明这一点. 尽管如此，在按平均值计算的确切意义下，这个极限的确存在.

8.6.1 定理 设 $A \in M_n$ 不可约非负矩阵，设 $Ax = \rho(A)x$，$A^Ty = \rho(A)y$，$x^Ty = 1$ 和 $L = xy^T$. 则

$$\lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} [\rho(A)^{-1}A]^m = L.$$

此外，存在有限正常数 $C = C(A)$ 使得

$$\left\| \frac{1}{N} \sum_{m=1}^{N} [\rho(A)^{-1}A]^m - L \right\| \leq \frac{C}{N}$$

对所有 $N = 1, 2, \cdots$ 成立.

证明：如果令 $\lambda=\rho(A)$，且选取 y 和 x 分别为 A 的左和右 Perron 向量，则引理(8.2.7)的假定(1)—(5)被满足，因而矩阵

$$I-[\rho(A)^{-1}A-L]=\rho(A)^{-1}[\rho(A)I-(A-\rho(A)L)]$$

524

是可逆矩阵. 利用引理(8.2.7)的(e)和本节末的习题1中的恒等式可算出

$$\begin{aligned}
&\frac{1}{N}\sum_{m=1}^{N}[\rho(A)^{-1}A]^m\\
&=\frac{1}{N}\sum_{m=1}^{N}([\rho(A)^{-1}A-L]^m+L)=L+\frac{1}{N}\sum_{m=1}^{N}[\rho(A)^{-1}A-L]^m\\
&=L+\frac{1}{N}\{\rho(A)^{-1}A-L\}\{I-[\rho(A)^{-1}A-L]^N\}\{I-[\rho(A)^{-1}A-L]\}^{-1}\\
&=L+\frac{1}{N}\{\rho(A)^{-1}A-L\}\{I-[\rho(A)^{-1}A]^N+L\}\{I-[\rho(A)^{-1}A-L]\}^{-1}.
\end{aligned}$$

在后一个表示式中与 N 有关的部分只有第二项的因子 $1/N$ 和项 $[\rho(A)^{-1}A]^N$，但是根据推论(8.1.33)，当 $N\to\infty$ 时，矩阵 $[\rho(A)^{-1}A]^N$ 各元一致有界. 因此，当 $N\to\infty$ 时，第二项的数量级为 $1/N$，因此它一致趋于零. □

分析一下引理(8.2.7)和(8.1.23)所需假设条件可知，相同的证法恰好证明下述更一般的(但叙述起来并不简短的)结果.

8.6.2 定理 设 $A\in M_n$ 是非负的，且设 x 和 y 是使 $Ax=\rho(A)x$ 和 $A^Ty=\rho(A)y$ 的非负向量. 如果

(a) $\rho(A)>0$；

(b) $x^Ty>0$；

(c) 矩阵

$$I-[\rho(A)^{-1}A-(x^Ty)^{-1}xy^T]$$

是可逆的；

(d) 当 $m\to\infty$ 时，$[\rho(A)^{-1}A]^m$ 一致有界；

则

$$\lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}[\rho(A)^{-1}A]^m=(x^Ty)^{-1}xy^T.$$

另外，存在有限正常数 $C=C(A)$，使得

$$\left\|\frac{1}{N}\sum_{m=1}^{N}[\rho(A)^{-1}A]^m-(x^Ty)^{-1}xy^T\right\|_\infty\leqslant\frac{C}{N}$$

对所有 $N=1，2，\cdots$ 成立.

525

习题

1. 如果 $B\in M_n$ 且 $I-B$ 可逆，证明

$$\sum_{m=1}^{N} B^m = B(I-B^N)(I-B)^{-1}.$$

提示：乘以 $I-B$.

2. 证明定理(8.6.2).

3. 试比较定理(8.5.1)和(8.6.1)的收敛速度．给出一个例子，说明(8.6.1)中的收敛速度不可能改进.

4. 假定 $A\in M_n$ 是不可约非负矩阵，且记 $A^m=[a_{ij}^{(m)}]$，$m=1，2，\cdots$．试用定理(8.6.1)证明，对每个给定的数对$(i，j)$，$a_{ij}^{(m)}>0$ 对 m 的无限多个值成立．这个结果可以看作定理(8.5.2)的推广．给出一个例子，说明也有 m 的无限多个值使得 $a_{ij}^{(m)}=0$.

5. 在定理(8.6.2)的假设条件下，证明，只要数对$(i，j)$使 $x_iy_j\neq 0$，则对 m 的无限多个值有 $a_{ij}^{(m)}>0$．这个结果为什么包括习题 4.

6. 当 A 是素矩阵时，直接证明定理(8.5.1)蕴涵定理(8.6.1)．**提示**：这里所需要的是证明从分析中得来的下述结果：如果序列收敛于有限极限，则它 Cesaro—可和于同一极限.

7. 假定 $A=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$．直接计算

$$\lim_{N\to\infty} N^{-1}\sum_{m=1}^{N}[\rho(A)^{-1}A]^m.$$

再计算用定理(8.6.1)给定的这个极限值，然后进行比较.

8.7 随机矩阵和双随机矩阵

如果非负矩阵 $A\in M_n$ 的所有行和是 1，则具有这一性质的矩阵称为(行)随机矩阵，这是因为每一行可以看成有 n 个点的样本空间上的离散概率分布．列随机矩阵是行随机矩阵的转置；这样的矩阵常常出现在(8.0)节中所讨论的城市间的人口流动模型中．随机矩阵还出现在 Markov 链的研究中以及象经济学和运筹学这类领域的各种各样的数学模型问题中.

526

M_n 中的随机矩阵的集合是紧凸集，并且具有简单而又重要的性质．如果用 $e\in\mathbf{R}^n$ 表示所有分量为 $+1$ 的向量，非负矩阵 $A\in M_n$ 是随机矩阵，当且仅当 $Ae=e$．所以，M_n 中的随机矩阵构成一个容易识别的非负矩阵族，且它们有一个特殊的正的公共特征向量．有正特征向量的非负矩阵具有许多特殊性质[例如，(8.1.30)，(8.1.31)和(8.1.33)]，因此所有随机矩阵具有这些性质.

如果随机矩阵 $A\in M_n$ 的转置 A^T 也是随机矩阵，就称 A 是双随机矩阵；所有行和与列和都是 $+1$．双随机矩阵的集合也是 M_n 中的紧凸集，显然，非负矩阵 $A\in M_n$ 是双随机矩阵，当且仅当 $Ae=e$ 且 $e^TA=e^T$．在(6.3.5)中已经遇到过一种双随机矩阵形式，即正交随机矩阵 $A=[|u_{ij}|^2]$，其中 $U=[u_{ij}]\in M_n$ 是酉矩阵．A 的行和与列和都是 $+1$ 可从 U 的行与列都是 Euclid 单位向量的事实推出.

双随机矩阵的另一个例子是置换矩阵组成的集合(群)．由 Birkhoff 定理可知，任一双随机矩阵是有限多个置换矩阵的凸组合，因而置换矩阵实际上是基本的和标准的双随机矩阵．我们

给出的关于 Birkhoff 定理的证明依赖于以下事实(见附录 B)：紧凸集 S 中的每个点是 S 的诸端点的凸组合. 我们将证明，双随机矩阵的集合的端点恰好是置换矩阵.

8.7.1 定理(Birkhoff)　矩阵 $A\in M_n$ 是双随机矩阵，当且仅当对某个 $N<\infty$，存在置换矩阵 $P_1,\cdots,P_N\in M_n$ 和正纯量 $\alpha_1,\cdots,\alpha_N\in\mathbf{R}$，使得 $\alpha_1+\cdots+\alpha_N=1$ 且 $A=\alpha_1P_1+\cdots+\alpha_NP_N$.

证明：条件的充分性是显然的；我们只需证明其必要性. 设 $A=[a_{ij}]\in M_n$ 是给定的双随机矩阵. 如果 A 是置换矩阵，则在它的每一行和每一列恰好有一个元素 $+1$，而所有其他元均为 0. 如果可以记 $A=\alpha_1B+\alpha_2C$，其中，$0<\alpha_1$，$\alpha_2<1$，$\alpha_1+\alpha_2=1$，且 B，C 是双随机矩阵，则因为 α_1 和 α_2 都是非零元且 b_{ij}、c_{ij} 是非负元，所以与 A 的 0 元 $a_{ij}=0$ 相对应的 B 和 C 的每个元一定适合 $0=a_{ij}=\alpha_1b_{ij}+\alpha_2c_{ij}$，因而 $b_{ij}=c_{ij}=0$. 由于 B 和 C 是双随机矩阵，所以它们的各个行和是 $+1$，因而各非零元一定都是 $+1$，且与 A 的各非零元处在相同的位置；即 $A=B=C$. 这就证明了每个置换矩阵是双随机矩阵集合的端点.

527

另一方面，如果 A 不是置换矩阵，则至少有 A 的一行，例如第 i 行，它至少含两个非零元. 在该行中选取任一非零元 $a_{i_1i_2}$，因为在第 i 行中至少有两个非零元且该行中的所有(非负)元之和是 $+1$，所以 $a_{i_1i_2}$ 一定适合 $0<a_{i_1i_2}<1$. 由于 $0<a_{i_1i_2}<1$，且第 i_2 列的所有(非零)元之和是 $+1$，所以在 $a_{i_1i_2}$ 所在的同一列，一定有另一个非零元 $a_{i_3i_2}$，$i_3\neq i_1$，且 $0<a_{i_3i_2}<1$. 根据同样的推理，在 $a_{i_3i_2}$ 所在的同一行，另有一个非零元 $a_{i_3i_4}$，$i_4\neq i_2$，且 $0<a_{i_3i_4}<1$. 如果把这个过程继续下去，且对用这种方式依次选出的各元加上标号，则经有限多步，原来选定的元素 a_{ij} 会首次再被选到. 从第一次出现元素 a_{ij} 直到第二次再出现 a_{ij} 的过程中，诸元素组成的序列(包括第一次出现的元 a_{ij}，但不包括第二次出现的元 a_{ij})是 A 的元素的有限有序序列，且其中每对相邻元在同一行或同一列交替出现；设 $a_{i'j'}$ 是这个序列中的最小(正)元. 设 $B\in M_n$ 是这样一个矩阵，在 B 中，$+1$ 出现在与序列的第一个元 a_{ij} 相对应的位置，-1 出现在与序列的第二个元相对应的位置，$+1$ 出现在与序列的第三个元相对应的位置，依此类推，交替地选取 ±1. B 的所有其他元是 0. 注意，B 的所有行和与列和是 0. 设 $A_+=A+a_{i'j'}B$ 及 $A_-=A-a_{i'j'}B$. 注意，A_+ 与 A_- 都是非负矩阵(因为 $a_{i'j'}$ 的极小性)，且它的行和与列和是 $+1$(因为 B 的行和与列和是 0)，所以 A_+ 和 A_- 是双随机矩阵. 由于 $A=\frac{1}{2}A_++\frac{1}{2}A_-$，且 $A_+\neq A$，我们得出 A 不是双随机矩阵集合的端点.

刚才给出的证明说明，一个给定的矩阵是双随机矩阵的紧凸集的端点，当且仅当它是置换矩阵. 从紧凸集的每个点是诸端点的凸组合的事实可知定理成立. □

因为在 M_n 中恰好有 $n!$ 个互不相同的置换矩阵，Birkhoff 定理保证，任意双随机矩阵可以表示成至多 $N=n!$ 个置换矩阵的凸组合. 更细致的分析说明，所需置换矩阵不超过 $N=n^2-2n+2$ 个.

习题

1. 设 $A\in M_n$ 是非零非负矩阵，且有正特征向量 $x=[x_i]$，又设 $D=\mathrm{diag}(x_1,\cdots,x_n)$，证明 $\rho=\rho(A)>0$，由(8.1.30)可知 $Ax=\rho x$，且 $ADe=\rho De$，其中 $e\in\mathbf{R}^n$ 的所有元是 $+1$，由此得出，A(经具有正主对角元的对角相似矩阵)相似于随机矩阵的正倍数[即 $\rho(A)$]. 这个论断使我

528

们把许多有关具有正特征向量的非负矩阵的问题化简成有关随机矩阵的问题.

2. 证明 M_n 中的随机矩阵的集合与双随机矩阵的集合是紧凸集.

3. 证明 M_n 中随机矩阵的集合与双随机矩阵的集合各自在矩阵乘法下构成一个半群；即如果 A，$B\in M_n$ 是(双)随机矩阵，则 AB 是(双)随机矩阵.

4. 证明，非负矩阵 $A\in M_n$ 是随机矩阵，当且仅当 $Ae=e$.

5. 证明 2×2 双随机矩阵是有相同主对角元的对称矩阵.

6. (a)试用(8.7.1)的证明中所采用的想法但不用附录 B 的结果，给出(8.7.1)的另一个直接证明，(b)然后给出一个算法，把一个双随机矩阵分解为置换矩阵的一个凸组合. **提示**：如果 A 不是置换矩阵，利用证明中所表述的诸元素组成的序列作一个置换矩阵，从 A 减去这个置换矩阵的正数倍就得到一个有相同行和与列和的非负矩阵，且它至少有一个非零元. 于是，在这个矩阵上重复上述过程且继续做下去.

7. 证明(8.7.1)中的分解是不唯一的.

8. 如果双随机矩阵 A 是可约的，证明 A 实际上置换相似于形如$\begin{bmatrix} A_2 & 0 \\ 0 & A_1 \end{bmatrix}$的矩阵. 关于 A_1 和 A_2 能说些什么?

进一步阅读 定理(8.7.1)的证明思想取自 B. Saunders and H. Schneider, "Applications of the Gordon-Stiemke Theorem in Combinatorial Matrix Theory," *SIAM Rev.* 21(1979), 528 541，从中可找到有关的事实. 每个双随机矩阵 $A\in M_n$ 至少是 n^2-2n+2 置换矩阵的一个凸组合，该结果的讨论可参看 M. Marcus and R. Ree, "Diagonals of Doubly Stochastic Matrices," *Quart. J. Math Oxford*, Ser. 2, 10(1959), 295-302.

529

附录A 复　　数

一个复数有形式

$$z = a + ib,$$

其中，a 和 b 是实数，而 i 是满足关系 $i^2 = -1$ 的形式符号．实数 a 称为 z 的实部，记作 $\mathrm{Re}\, z$；实数 b 称为 z 的虚部，记作 $\mathrm{Im}\, z$．复数 $z=a+ib$ 的复共轭 $\bar{z}$ 是 $\bar{z}=a-ib$．如果 $z_1=a_1+ib_1$ 和 $z_2=a_2+ib_2$ 是复数，则其加法和乘法的二元运算通过实数的相应运算自然定义为如下形式：

$$z_1 + z_2 = (a_1 + a_2) + i(b_1 + b_2), \quad z_1 z_2 = a_1 a_2 - b_1 b_2 + i(a_1 b_2 + a_2 b_1).$$

因此，加法是实部和虚部分别相加的结果，而乘法是代数展开及关系 $i^2=-1$ 的结果．$z=a+ib$ 的加法逆是 $-z=-a+i(-b)$．而当 $z\neq 0=0+i0$ 时，z 的乘法逆是

$$\frac{1}{z} = \frac{a-ib}{a^2+b^2} = \frac{a}{a^2+b^2} + i\left(\frac{-b}{a^2+b^2}\right).$$

复数 z_1 和 z_2 的减法和除法定义为

$$z_1 - z_2 = z_1 + (-z_2), \quad \frac{z_1}{z_2} = z_1\left(\frac{1}{z_2}\right) = \frac{z_1\bar{z}_2}{z_2\bar{z}_2}.$$

所有复数的集合记作 $\mathbf{C}$；加法和乘法运算是可交换的，并且 $\mathbf{C}$ 在这些运算下构成一个域，以实数 $0=0+i0$ 为加法单位元，以实数 $1=1+i0$ 为乘法单位元．实数域 $\mathbf{R}$ 构成 $\mathbf{C}$ 的一个子域；z 的绝对值(或模)，记作 $|z|$，定义为 $|z|=+(z\bar{z})^{1/2}$，它总是一个非负实数．如果 $z_2\neq 0$，则商 z_1/z_2 为 $(1/|z_2|^2)z_1\bar{z}_2$．容易验证，乘法运算与复共轭是可交换的，$\overline{z_1 z_2}=\bar{z}_1\bar{z}_2$，且复共轭的复共轭又是原来的复数．因为 $\mathrm{Re}\, z=(1/2)(z+\bar{z})$ 和 $\mathrm{Im}\, z=(1/2i)(z-\bar{z})$，所以实数就是使 $\mathrm{Im}\, z=0$ 的那样一些 $z\in\mathbf{C}$，或等价地有 $z=\bar{z}(=\mathrm{Re}\, z)$．

531

几何上，复数域 $\mathbf{C}$ 可以看作一个具有原点 0，一条"实轴"和一条"虚轴"的平面．因此，$z=a+ib$ 可等同于有序偶 (a,b)．实轴 $\{z:\mathrm{Im}\, z=0\}$ 就是通常的实线，而虚轴 $\{z:\mathrm{Re}\, z=0\}$ 就是实线的 i 倍或所有"纯虚"数．$z\in\mathbf{C}$ 到实轴(虚轴)的投影是 $\mathrm{Re}\, z$ ($i\mathrm{Im}\, z$)．复共轭是经实轴的反射．且 $|z|$ 是复平面中原点到 z 的 Euclid 距离．$\mathbf{C}$ 的开(闭)右半平面是 $\{z\in\mathbf{C}:\mathrm{Re}\, z>(\geqslant)0\}$．而 $\mathbf{C}$ 的开(闭)上半平面是 $\{z\in\mathbf{C}:\mathrm{Im}\, z>(\geqslant)0\}$．$\mathbf{C}$ 的单位圆盘是 $\{z\in\mathbf{C}:|z|\leqslant 1\}$，而以 r 为半径，以 $a\in\mathbf{C}$ 为中心的圆盘是 $\{z\in\mathbf{C}:|z-a|\leqslant r\}$．

上一段用直角坐标描述了复平面 $\mathbf{C}$．同时复平面也可以有效地用极坐标来描述，$z\in\mathbf{C}$ 在平面的位置由 r 和 θ 所确定，其中 r 是 z 所在的以原点为中心的圆的半径，而 θ 是按反时针方向从实线正向转到 z 所在的从原点出发的有向射线的角度．于是 z 的极坐标为 (r,θ)，并且通常用记号 $z=re^{i\theta}$ 来表示，其中 $e^{i\theta}\equiv\cos\theta+i\sin\theta$．因此，如果 $z=a+ib$ 为直角坐标，且 $z=re^{i\theta}$ 为极坐标，则从极坐标到直角坐标的变换是

$$a = r\cos\theta, \quad b = r\sin\theta,$$

而从直角坐标到极坐标的变换(当 $r\neq 0$ 时)是

$$r=|z|=(a^2+b^2)^{1/2},\quad \theta=\arcsin\frac{b}{r}=\arg z,$$

这里，我们一般取 $0\leqslant\theta<2\pi$. 圆形对象常常容易用极坐标来描述. 例如，C 中的单位圆盘是 $\{re^{i\theta}: 0\leqslant r\leqslant 1,\ 0\leqslant\theta<2\pi\}$.

532

附录B 凸集和凸函数

设 V 是含有实数的一个域上的向量空间. V 的一组选定的元素 v_1，…，$v_k\in V$ 的一个凸组合是其系数非负且和为 1 的线性组合：

$$\alpha_1 v_1+\cdots+\alpha_k v_k;\quad \alpha_1,\cdots,\alpha_k\geqslant 0,\sum_{i=1}^{k}\alpha_i=1.$$

V 的子集 K 称为凸集，是指从 K 中任意选定一组元素所作成的任一凸组合仍在 K 中. 等价地，K 是凸集，是指 K 中各个点偶的所有凸组合仍在 K 中. 几何上，这可以解释为连结 K 的任意两点的线段必定在 K 中；即 K 没有"凹"或"洞". 一个凸集 K 称为凸锥，是指每当 $\alpha>0$ 和 $x\in K$，就有 $\alpha x\in K$(等价地，K 中元素的正线性组合仍在 K 中). 直接验证可知，两个凸集(相应地，凸锥)的集和及交仍为一个凸集(相应地，凸锥).

一个闭凸集 K 的一个端点是这样一个点 $z\in K$，它只能用一种平凡的方式写成 K 中点的一个凸组合，也就是说，$z=\alpha x+(1-\alpha)y$，$0<\alpha<1$，x，$y\in K$，蕴涵 $x=y=z$. 一个闭凸集可能有有限个端点(例如，一个多面体)，也可能有无限多个端点(例如，一个闭圆盘)，还可能无端点(例如，$\mathbf{R}^2$ 中的闭上半平面). 然而，一个紧凸集总有端点. V 中点的一个集合 S 的凸包，记作 Co(S)，就是选择 S 的所有点组作成的所有凸组合的集合. 或等价地，它是包含 S 的(所有凸集的交)最小凸集. Krein-Milman 定理是说，一个紧凸集是它的端点的凸包，我们说一个紧凸集是有限生成的，是指它有有限多个端点，这些端点称为该凸集的生成元.

现在假定 V 是具有一个给定范数的实或复向量空间，分离超平面定理是说，对于两个不相交的凸集 $K_1\subseteq V$ 和 $K_2\subseteq V$，存在 V 中的一个超平面 H，使得 K_1 位于由 H 确定的两个闭半空间中的一个内，而 K_2 位于另一个内；即 H 分离 K_1 和 K_2. V 中的一个超平面 H 正好是 V 的一个一维子空间的正交补的一个平移：$H=\{x\in V:\langle x-p,\ q\rangle=0\}$，其中，$q\in V$ 是给定的向量，$q\neq 0$. 超平面 H 确定两个开半空间：$H^+=\{x\in V:\langle x-p,\ q\rangle>0\}$，$H^-=\{x\in V:\langle x-p,\ q\rangle<0\}$. 集合 $H_0^+=H^+\cup H$ 和 $H_0^-=H^-\cup H$ 是由 H 所确定的两个闭半空间. 因此，分离意味着对某两个向量 p，q 有 $K_1\subseteq H_0^+$ 和 $K_2\subseteq H_0^-$. 对两个凸集附加某些条件就会有各种强化的分离结论. 例如，如果 K_1 和 K_2 的闭包不相交，那么分离可以取得更为精确；即 $K_1\subseteq H^+$，$K_2\subseteq H^-$. 任一有界集 $S\subset V$ 的凸包的闭包可以通过取包含 S 的所有闭半空间的交而得到.

在 V 是具有复内积 $\langle\cdot,\ \cdot\rangle$ 的向量空间 $\mathbf{C}^n$ 的情形，也可以类似地定义超平面和半空间，只是必须把 $\mathbf{C}^n$ 和 $\mathbf{R}^{2n}$ 等同起来，且必须用下述实内积 Re$\langle\cdot,\ \cdot\rangle$ 代替 $\langle\cdot,\ \cdot\rangle$. 将 $x+iy\in\mathbf{C}^n$ 等同于 $\begin{bmatrix}x\\y\end{bmatrix}\in\mathbf{R}^{2n}$，还应注意到，根据复内积的共轭线性性质，$\mathrm{Re}\langle x_1+iy_1,\ x_2+iy_2\rangle=\langle x_1,\ x_2\rangle+\langle y_1,\ y_2\rangle$. 于是 $\langle x_1,\ x_2\rangle+\langle y_1,\ y_2\rangle$ 是 $\begin{bmatrix}x_1\\y_1\end{bmatrix}$ 和 $\begin{bmatrix}x_2\\y_2\end{bmatrix}$ 的(实)内积，而且 $\mathbf{R}^{2n}$ 中定义的超平面和半空间在 $\mathbf{C}^n$ 中有相应的几何解释.

我们说定义在一个凸集 $K\subseteq V$ 上的实值函数 f 是凸函数，指的是对所有 $0<\alpha<1$ 和所有 $x,y\in K$，$y\neq x$，有

$$f(\alpha x+(1-\alpha)y)\leqslant \alpha f(x)+(1-\alpha)f(y). \qquad (*)$$

如果上述不等式总是严格的，那么 f 称为严格凸函数．如果对所有 $0<\alpha<1$ 和所有 x，$y\in K$，$y\neq x$，上述不等式是反向的，那么 f 称为凹函数(或严格凹函数，如果反向不等式总是严格的)．等价地，一个凹函数(相应地，严格凹函数)正是一个凸函数(相应地，严格凸函数)的负值．几何上，连结任意两个函数值的弦位于一个凸函数(相应地，凹函数)的图象的上方(相应地，下方)．一个线性函数既是凸的，也是凹的．在 $V=\mathbf{R}^n$ 和 K 是开集的情形，Hessian 矩阵为

534

$$H(x)\equiv\left[\frac{\partial^2 f}{\partial x_i\,\partial x_j}(x)\right],$$

它是 $M_n(\mathbf{R})$中的对称矩阵，对于一个凸函数 f，$H(x)$在 K 中几乎处处存在，并且在使 $H(x)$ 存在的 K 中的各个点，它一定是半正定的．在严格凸的情形，它是正定的．反过来，一个函数，如果它的 Hessian 矩阵在整个凸集上是半正定(相应地，正定)的，那么该函数是凸(相应地，严格凸)的．类似地，负定性对应凹性．

凸函数和凹函数的最优化有一些好的性质．在紧凸集上，凸函数(相应地，凹函数)的极大值(相应地，极小值)在一个端点达到．另一方面，在一个凸集上，由那些达到一个凸函数(相应地，凹函数)的极小值(相应地，极大值)的点组成的集合是凸的，并且任一局部极小值(相应地，极大值)是全局极小值(相应地，极大值)．例如，一个严格凸函数至多在凸集的一个点达到极小值．且一个临界点一定是极小值点．

实数的凸组合满足某些简单而又常用的不等式．如果 x_1，…，x_k 是给定的实数，那么

$$\min_{1\leqslant i\leqslant k} x_i\leqslant\sum_{i=1}^{k}\alpha_i x_i\leqslant\max_{1\leqslant i\leqslant k} x_i$$

对任意凸组合 α_1，α_2，…，$\alpha_k\geqslant 0$ 和 $\alpha_1+\cdots+\alpha_k=1$ 成立．

研究定义在一个区间上的某些简单的单变量凸函数 $f(\cdot)$可诱导出各种各样的经典不等式．我们可以用归纳法证明，定义在区间上的关于两个点的不等式(*)可以推出关于 n 个点的不等式

$$f\Big(\sum_{i=1}^{n}\alpha_i x_i\Big)\leqslant\sum_{i=1}^{n}\alpha_i f(x_i),\quad n=2,3,\cdots, \qquad (**)$$

只要 $\alpha_i\geqslant 0$，$\alpha_1+\cdots+\alpha_n=1$，且所有 x_i 都在区间内．

应用(**)于区间$(0,\infty)$上的严格凸函数 $f(x)=-\log x$，便导出加权算术—几何平均值不等式

$$\sum_{i=1}^{n}\alpha_i x_i\geqslant\prod_{i=1}^{n}x_i^{\alpha_i},\quad x_i\geqslant 0.$$

当所有 $\alpha_i=1/n$ 时，它就是算术—几何平均值不等式

$$\frac{1}{n}\sum_{i=1}^{n}x_i\geqslant\Big(\prod_{i=1}^{n}x_i\Big)^{1/n},\quad x_i\geqslant 0,$$

等式成立，当且仅当所有 x_i 都相等. [535]

把(* *)应用于定义在区间(0, ∞)上的 $f(x)=x^p$，$p>1$，便导出 Hölder 不等式

$$\sum_{i=1}^{n} x_i y_i \leqslant \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left(\sum_{i=1}^{n} y_i^q\right)^{1/q},$$

其中 x_i，$y_i>0$，$p>1$，且 $1/p+1/q=1$. 等式成立，当且仅当向量$[x_i^p]$和$[y_i^q]$相关. 如果取 $p=q=2$，我们便得到 Cauchy-Schwarz 不等式的另一种形式

$$\sum_{i=1}^{n} x_i y_i \leqslant \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \left(\sum_{i=1}^{n} y_i^2\right)^{1/2}.$$

等式成立，当且仅当向量$[x_i]$和$[y_i]$是相关的. 我们可以从 Hölder 不等式推出 Minkowski 不等式

$$\left[\sum_{i=1}^{n} (x_i+y_i)^p\right]^{1/p} \leqslant \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p},$$

其中 x_i，$y_i>0$ 且 $p\geqslant 1$. 等式成立，当且仅当向量$[x_i]$和$[y_i]$是相关的.

进一步阅读 关于凸集与几何的其他资料可参看[Val]. 关于凸函数与不等式的其他资料可参看[Boa]和[BB]. [536]

附录C 代数基本定理

促使人们去建立复数域C的一个历史动机是实系数多项式可能有非实的复零点．例如，二次公式表明，方程 $x^2-2x+2=0$ 有根(解)$\{1+i, 1-i\}$．然而，任一实系数多项式的所有零点都包含在C中．事实上，当把可能的系数域扩充到C时，所有复系数多项式的所有零点仍然包含在C中．因此，C是代数闭域的一个例子；也就是，不存在以C为其子域的域F，使得有一个系数取自C的多项式，它的根在F中而不在C中．

代数基本定理是说，次数至少是1的任一复系数多项式在复数中至少有一个零点z[即z是方程$p(x)=0$的一个根]．运用综合除法可知，如果z是$p(x)$的零点，那么$x-z$就除尽$p(x)$；即，$p(x)=(x-z)q(x)$，其中，$q(x)$是复系数多项式，它的次数比$p(x)$少1．于是$p(x)$的诸零点是$q(x)$的零点及z．

下述定理是代数基本定理的一个推论．

定理 每个次数为$n\geq 1$的复系数多项式在复数域中恰好有n个根(重根按重数计算)．

$p(x)=0$的一个根的重数是使$(x-z)^k$除尽$p(x)$的最大整数k，即z作为$p(x)=0$的根出现的次"数"．如果一个根有重数3，那么在$p(x)=0$的n个根中，它要累计3次．由此推出，每个复系数多项式在复数域上总可以分解成线性因式的乘积．

537

不过，如果实系数多项式$p(x)$有某些非实复根，那么它们必须成共轭对出现，因为，如果$0=p(z)$，则$0=\overline{0}=\overline{p(z)}=p(\overline{z})$．由于

$$(x-z)(x-\overline{z}) = x^2 - 2\mathrm{Re}(z)x + |z|^2,$$

由此得出，任一实多项式在实数域上可分解成线性因式和二次因式的幂的乘积，而每个不可约二次因式对应一对共轭复根．

538

进一步阅读 关于代数基本定理的一个初等证明可参看[Chi]．

附录 D　多项式的零点对其系数的连续依赖性

运用复分析不难证明一个重要的事实，那就是，一个次数为 $n\geqslant 1$ 的复系数多项式的 n 个零点连续地依赖于它的系数.

对于 $x\in\mathbf{C}^n$，设 $f(x)=[f_1(x),\cdots,f_m(x)]^T$，其中 $f_i:\mathbf{C}^n\to\mathbf{C}$，$i=1,\cdots,m$. 函数 $f:\mathbf{C}^n\to\mathbf{C}^m$ 在 x 连续，是指每个 f_i 在 x 连续，$i=1,\cdots,m$. 函数 $f_i:\mathbf{C}^n\to\mathbf{C}$ 在 x 点连续，是指对每个 $\varepsilon>0$，都存在 $\delta>0$，使得只要 $\|y-x\|<\delta$ 就有 $|f_i(y)-f_i(x)|<\varepsilon$，其中 $\|\cdot\|$ 是 $\mathbf{C}^n$ 上的向量范数.

可以用下面的说法来直观地叙述连续依赖的结论：把首系数为 1 的 n 次多项式的 n 个系数(首系数 1 除外)与该多项的 n 个零点对应起来的函数 $f:\mathbf{C}^n\to\mathbf{C}^n$ 是连续的. 可是这里有个难以解决的问题，不存在定义这个函数的简单方式，这是因为，在 n 个零点中间不存在定义一个顺序的自然方式. 我们提出下面的定理，作为多项式的零点对其系数的连续依赖的一个精确叙述.

定理　设 $n\geqslant 1$，且设

$$p(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0,\quad a_n\neq 0$$

是一个复系数多项式，那么，对任给 $\varepsilon>0$，存在 $\delta>0$，使得对任一适合 $b_n\neq 0$ 的多项式

$$q(x)=b_nx^n+b_{n-1}x^{n-1}+\cdots+b_1x+b_0$$

539

和

$$\max_{0\leqslant i\leqslant n}|a_i-b_i|<\delta,$$

我们有

$$\min_{\tau}\max_{1\leqslant j\leqslant n}|\lambda_j-\mu_{\tau(j)}|<\varepsilon,$$

其中，$\lambda_1,\cdots,\lambda_n$ 是 $p(x)$ 的零点，而 $\mu_1,\cdots,\mu_n$ 是按某个顺序排列的 $q(x)$ 的零点，且(极小) min 取遍 $1,2,\cdots,n$ 的所有排列 τ.

因此，一个多项式的各个系数的足够小的变化只会引起任一零点的一个小的变化，这一原理在矩阵分析中有非常重要的意义，这是因为，矩阵 $A\in M$ 的特征多项式 $p_A(t)$ 的诸系数是 A 的各元的连续函数(实际上是多项式)(1.2.11)，且 $p_A(t)$ 的零点是 A 的特征值. 因为连续函数的复合是连续的，A 的各元的足够小的变化只会引起 $p_A(t)$ 的系数的一个小的变化，而它只会导致特征值的一个小变化. 因此，实的或复的方阵的特征值连续地依赖于它的各个元.

进一步阅读　关于 $p(x)$ 和 $q(x)$ 的各零点间之差 ε 的显式界可用其系数分离 δ 及其系数的大小来表示的问题可参看 L. Elsner, "On the Variation of the Spectra of Matrices," *Linear Algebra Appl.* 47(1982), 127-138.

540

附录 E Weierstrass 定理

设 V 是具有范数 $\|\cdot\|$ 的有限维实或复向量空间. 关于中心为 $x\in V$, 半径为 ε 的球是 $B_\varepsilon(x)=\{y\in V: \|y-x\|\leqslant\varepsilon\}$. 我们称子集 $S\subseteq V$ 是开集, 是指对每个 $x\in S$, 都存在一个 $\varepsilon>0$, 使得 $B_\varepsilon(x)\subseteq S$. 子集 $T\subseteq V$ 称闭集, 是指 T 在 V 中的补集是开集. 子集 $S\subseteq V$ 称为有界的, 是指存在 $r>0$, 使得 $S\subseteq B_r(0)$. 等价地, T 是闭的, 当且仅当 T 的任一收敛序列(关于 $\|\cdot\|$)的极限都在 T 中, 而 S 是有界的, 是指 S 包含在具有有限半径的任一球中. 子集 $S\subseteq V$ 是紧集, 是指它既是闭的, 又是有界的.

对于 $S\subseteq V$, 一个函数 $f: S\to\mathbf{R}$ 在 S 上可以达到或者不可以达到一个(全局)极大值或极小值. 但是, 在某些常见的情形, 我们可以确信, f 在 S 上达到一个极大值.

定理(Weierstrass) 设 S 是有限维实或复向量空间 V 的紧子集. 如果 $f: S\to\mathbf{R}$ 是连续函数, 那么存在点 $x_{\min}\in S$, 使得对所有 $x\in S$ 有

$$f(x_{\min})\leqslant f(x),$$

且存在点 $x_{\max}\in S$, 使得对所有 $x\in S$ 有

$$f(x)\leqslant f(x_{\max}).$$

即 f 在 S 上达到它的极小值和极大值. 当然, 可能不止在 S 的一个点上达到值 $\max_{x\in S}f(x)$ 和 $\min_{x\in S}f(x)$.

[541] 如果 Weierstrass 定理的两个主要假定(紧的 S 和连续的 f)不成立, 结论可能不真. 但是, S 是一个有限维实或复向量空间的子集不是本质的. 对子紧集的一个适当的定义, Weierstrass 定理对定义在一般拓扑空间的一个紧子集上的连续实值函数成立. [542]

参考文献

Ait A. C. Aitken. *Determinants and Matrices*. 9th ed. Oliver and Boyd, Edinburgh, 1956.

Bar 75 S. Barnett. *Introduction to Mathematical Control Theory*. Clarendon Press, Oxford, 1975.

Bar 79 S. Barnett. *Matrix Methods for Engineers and Scientists*. McGraw-Hill, London, 1979.

Bar 83 S. Barnett. *Polynomials and Linear Control Systems*. Dekker, New York, 1983.

BB E. F. Beckenbach and R. Bellman. *Inequalities*. Springer-Verlag, New York, 1965.

Bel R. Bellman. *Introduction to Matrix Analysis*. 2d ed. McGraw-Hill, New York, 1970.

Boa R. P. Boas, Jr. *A Primer of Real Functions*. 2d ed. Carus Mathematical Monographs, No. 13. Mathematical Association of America, Washington, D.C., 1972.

BPl A. Berman and R. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, New York, 1979.

BSt S. Barnett and C. Storey. *Matrix Methods in Stability Theory*. Barnes & Noble, New York, 1970.

CaLe J. A. Carpenter and R. A. Lewis. *KWIC Index for Numerical Algebra*. U.S. Dept. of Commerce, Springfield, Va. Microfiche and printed versions available from National Technical Information Service, U.S. Dept. of Commerce, 5285 Port Royal Road, Springfield, VA 22161.

Chi L. Childs. *A Concrete Introduction to Higher Algebra*. Springer-Verlag, Berlin, 1979.

Cul C. G. Cullen. *Matrices and Linear Transformations*. Addison-Wesley, Reading, Mass., 1966.

Don W. F. Donoghue, Jr. *Monotone Matrix Functions and Analytic Continuation*. Springer-Verlag, Berlin, 1974.

Fad V. N. Faddeeva. Trans. C. D. Benster. *Computational Methods of Linear Algebra*. Dover, New York, 1959.

Fan Ky Fan. *Convex Sets and Their Applications*. Lecture Notes, Applied Mathematics Division, Argonne National Laboratory, Summer 1959.

Fie M. Fieldler. *Spectral Properties of Some Classes of Matrices*. Lecture Notes, Report No. 75.01R. Chalmers University of Technology and the University of Göteborg, 1975.

Fra J. Franklin. *Matrix Theory*. Prentice-Hall, Englewood Cliffs, N.J., 1968.

Gan F. R. Gantmacher. *The Theory of Matrices*. 2 vols. Chelsea, New York, 1959.

Gant F. R. Gantmacher. *Applications of the Theory of Matrices*. Interscience, New York, 1959.

GKr F. R. Gantmacher and M. G. Krein. *Oszillationsmatrizen, Oszillationskerne, und kleine Schwingungen mechanische Systeme.* Akademie-Verlag, Berlin, 1960.

GLR 82 I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials.* Academic Press, New York, 1982.

GLR 83 I. Gohberg, P. Lancaster, and L. Rodman. *Matrices and Indefinite Scalar Products.* Birkhäuser-Verlag, Boston, 1983.

Grah A. Graham. *Kronecker Products and Matrix Calculus with Applications.* Horwood, Chichester, U.K., 1981.

Gray F. A. Graybill. *Matrices with Applications to Statistics.* 2d ed. Wadsworth, Belmont, Calif., 1983.

Gre W. H. Greub. *Multilinear Algebra.* 2d ed. Springer-Verlag, New York, 1978.

GVl G. Golub and C. VanLoan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, 1983.

Hal 58 P. R. Halmos. *Finite-Dimensional Vector Spaces.* Van Nostrand, Princeton, N.J., 1958.

Hal 67 P. R. Halmos. *A Hilbert Space Problem Book.* Van Nostrand, Princeton, N.J., 1967.

HJ R. Horn and C. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, 1989.

HKu K. Hoffman and R. Kunze. *Linear Algebra.* 2d ed. Prentice-Hall, Englewood Cliffs, N.J., 1971.

Hou 64 A. S. Householder. *The Theory of Matrices in Numerical Analysis.* Blaisdell, New York, 1964.

Hou 72 A. S. Householder. *Lectures on Numerical Algebra.* Mathematical Association of America, Buffalo, N.Y., 1972.

HSm M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press, New York, 1974.

Jac N. Jacobson. *The Theory of Rings.* American Mathematical Society, New York, 1943.

Kap I. Kaplansky. *Linear Algebra and Geometry: A Second Course.* Allyn & Bacon, Boston, 1969.

Kar S. Karlin. *Total Positivity.* Stanford University Press, Stanford, Calif., 1960.

Kel R. B. Kellogg. *Topics in Matrix Theory.* Lecture Notes, Report No. 71.04, Chalmers Institute of Technology and the University of Göteberg, 1971.

Kow H. Kowalsky. *Lineare Algebra.* 4th ed. deGruyter, Berlin, 1969.

LaH C. Lawson and R. Hanson. *Solving Least Squares Problems.* Prentice-Hall, Englewood Cliffs, N.J., 1974.

Lan P. Lancaster. *Theory of Matrices.* Academic Press, New York, 1969.

LaTi P. Lancaster and M. Tismenetsky. *The Theory of Matrices With Applications.* 2d ed. Academic Press, New York, 1985.

Mac C. C. MacDuffee. *The Theory of Matrices.* Chelsea, New York, 1946.

Mar M. Marcus. *Finite Dimensional Multilinear Algebra.* 2 vols. Dekker, New York, 1973–75.

Mir L. Mirsky. *An Introduction to Linear Algebra.* Clarendon Press, Oxford, 1963.

MMi M. Marcus and H. Minc. *A Survey of Matrix Theory and Matrix Inequalities.* Allyn & Bacon, Boston, 1964.

MOl A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York, 1979.

Mui T. Muir. *The Theory of Determinants in the Historical Order of Development.* 4 vols. MacMillan, London, 1906, 1911, 1920, 1923; Dover, New York, 1966. *Contributions to the History of Determinants, 1900–1920.* Blackie, London, 1930.

Ner E. Nering. *Linear Algebra and Matrix Theory.* 2d ed. Wiley, New York, 1963.

New M. Newman. *Integral Matrices.* Academic Press, New York, 1972.

Nob B. Noble. *Applied Linear Algebra.* Prentice-Hall, Englewood Cliffs, N.J., 1969.

Per S. Perlis. *Theory of Matrices.* Addison-Wesley, Reading, Mass., 1952.

Rog G. S. Rogers. *Matrix Derivatives.* Lecture Notes in Statistics, Vol. 2. Dekker, New York, 1980.

Rud W. Rudin. *Principles of Mathematical Analysis.* 3rd ed. McGraw-Hill, New York, 1976.

Sen E. Seneta. *Nonnegative Matrices.* Wiley, New York, 1973.

Ste G. W. Stewart. *Introduction to Matrix Computations.* Academic Press, New York, 1973.

Str G. Strang. *Linear Algebra and Its Applications.* Academic Press, New York, 1976.

STy D. A. Suprenenko and R. I. Tyshkevich. *Commutative Matrices.* Academic Press, New York, 1968.

Tod J. Todd (ed.). *Survey of Numerical Analysis.* McGraw-Hill, New York, 1962.

TuA H. W. Turnbull and A. C. Aitken. *An Introduction to the Theory of Canonical Matrices.* Blackie, London, 1932.

Tur H. W. Turnbull. *The Theory of Determinants, Matrices and Invariants.* Blackie, London, 1950.

Val F. A. Valentine. *Convex Sets.* McGraw-Hill, New York, 1964.

Var R. S. Varga. *Matrix Iterative Analysis.* Prentice-Hall, Englewood Cliffs, N.J., 1962.

Wed J. H. M. Wedderburn. *Lectures on Matrices.* American Mathematical Society Colloquium Publications XVII. American Mathematical Society, New York, 1934.

Wie H. Wielandt. *Topics in the Analytic Theory of Matrices.* Lecture Notes prepared by R. Meyer. Department of Mathematics, University of Wisconsin, Madison, 1967.

Wil J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, 1965.

索　引

索引中的页码为英文原书页码，与书中页边标注的页码一致.

A

a priori bounds （先验界），337
absolute （绝对）
  convergence （收敛），279，300
  value of a complex number （复数的值），532
vector norm （向量范数），285，310，365，438
additive property，of inner product （内积的可加性），260
adjoint （伴随）
  classical （经典），20
  Hermitian （Hermite），6
adjugate （转置伴随），20
algebraically （代数）
  closed field （闭域），41，537
alternating sum （交错和），8
angle between vectors （向量间夹角），15
annihilate （零化），142
antilinear transformation （反线性变换），250
approximation problems （逼近问题），332，427
augmented matrix （增广矩阵），11，12

B

back substitution （后向替换），159
ball of radius r （半径为 r 的球），281，541
basis （基），3
  change of （变换），30
  orthonormal （正交），16
  representation （表示），31
bilinear form （双线性型），169，175
biorthogonality （双正交性），59
Birkhoff's theorem （Birkhoff 定理），197，527
Bochner's theorem （Bochner 定理），394
boundary （边界），282
bounded set （有界集）282，541
Brauer
  condition for invertibility （可逆条件），381
  region （区域），380
  theorem （定理），380
Brualdi
  condition for invertibility （可逆性条件），389
  region （区域），385
  theorem （定理），385，387

C

cancellation theorem （消去定理），78，141
canonical forms （标准形）
  consimilarity （合相似），251
  integer matrices （整数矩阵），158
  irreducible normal form （不可约），506
  Jordan，121
  rational canonical （有理），156
  rational matrices （有理矩阵），158
  real Jordan （实 Jordan），152
  real orthogonal matrices （实正交矩阵），108
  real skew-symmetric matrices （实斜对称矩阵），107
  real symmetric matrices （实对称矩阵），107
  singular value decomposition （奇异值分解），157，414ff
  symmetric Jordan （对称 Jordan），209
  triangular factorization （三角分解），157
Carmichael and Mason's bound on zeroes （关于零点的 Carmichael 界），317，318，364
Cassini，ovals of （Cassini 椭圆形），380
Cauchy
  sequence （序列），274
  bound on zeros （关于零点的界），316，318，364
Cauchy Binet formula （Cauchy Binet 公式），22
Cayley-Hamilton theorem （Cayley-Hamilton 定理），86
characteristic equation （特征方程），87
characteristic polynomial （特征多项式），38，86，87，540

Cholesky factorization (Cholesky 分解), 114, 407
closed set (闭集), 282, 511
closure (闭包), 282
cofactor (代数余子式), 17
column rank (列秩), 12
commutative ring (交换环), 95
commutator (换位子), 98
commuting family (交换族), 51, 81, 99, 139
compact set (紧集), 282, 511
completeness property of a vector space (向量空间的完备性), 271
complex (复)
  conjugate (共轭), 531
  numbers (数), 531, 532
concave function (凹函数), 534
condiagonalization (合对角化), 244, 248
condition number (条件数), 336, 340, 365, 366, 374, 442
coneigenvalue (合特征值)
  characterization (特征), 246
  definition (定义), 245
coneigenvector (合特征向量), 245
conformal (共形的), 17
congruence (相合)
  * congruence (* 相合), 220, 399, 464ff, 470
  simultaneous * congruence, canonical pairs (同时 * 相合标准形偶), 236
  T congruence (T 相合), 225
conjugate linear (共轭线性), 169
conjunctive (共轭相合), 220
consimilarity (合相似), 234, 244
  characterizations (的特征), 251
  to a real matrix (于一个实矩阵), 255
consistent (相容)
  linear equations (线性方程组), 12
  vector norm (向量范数), 324
constrained extrema (约束极值), 34
continuous dependence of eigenvalues (特征值的连续依赖), 540
contriangularization (合三角化), 244
convergence of a sequence (序列的收敛), 269
convex (凸), 284
  combination (组合), 535
  cone (锥), 463
  function (函数), 392, 533, 534-536
  hull (包), 533
  sets (集), 533-536
coordinate representation (坐标表示), 30
Courant-Fischer theorem (Courant Fischer 定理), 179, 420, 424, 472
Cramer's rule (Cramer 法则), 21
cycle (回路), 357
cyclic of index k (指标 k 的循环), 514

D

defect from normality (正规性亏损值), 316
defective (亏损), 58
deflation (压缩), 63, 83
deleted absolute row sums (去心绝对行和), 344
dependent (相关), 3
determinant (行列式), 7, 11, 398
determinantal inequalities (行列式不等式), 453, 467, 476-486
  Fischer, 478
  Geršgorin, 351
  Hadamard, 477
  Hadamard-Fischer, 485
  Minkowski, 482
  Oppenheim, 480
  Ostrowski-Taussky, 481
  Szasz, 479
diagonalizable (可对角化), 139, 145
  by orthogonal similarity (用正交相似), 211
  definition (定义), 46
  orthogonality (正交), 101
  simultaneously (同时), 49
  unitary (酉), 101
diagonalization (对角化)
  by congruence (用相合), 228
  by consimilarity (用合相似), 234, 244, 248
  by similarity (用相似), 46, 145
  by unitary congruence (用酉相合), 204
  by unitary consimilarity (用酉合相似), 244, 245
  by unitary similarity (用酉相似), 101

simultaneous （同时），52
diagonally dominant （对角占优），349
strictly （严格），302，349
difference scheme （差分方法），394
differential equations （微分方程），132，394
elliptic （椭圆型），239，459
hyperbolic （双曲型），239
partial （偏），168，216，218
dimension （维数），4
direct sum （直和），24
directed （有向）
graph of a matrix （矩阵图），357，517，522
path （道路），357
dual pair （对偶对），278
duality theorem （对偶性定理），287

E

edges （边），168
eigenspace （特征空间），57
eigenvalue （特征值）
algebraic multiplicity （代数重数），58，60，138，141，497，499
algebraically simple （代数单重），371
definition （定义），35
deflation to calculate （计算的压缩），63
distinct （互异），48
dominant （优势），506
generalized （广义），213
geometric multiplicity （几何重数），58，60，138，141，497，498
ill conditioned （病态），367
inclusion region （包含区域），501
inclusion theorem （包含定理），177
location （估计），343
moments （矩），43
of a sum （和的），181，184
perfectly conditioned （优态），367
power method to calculate （求的幂法），62
principal submatrices （主子矩阵的），189
well-conditioned （良态），367
eigenvector （特征向量），57
definition （定义），35
left （左），59，371
positive （正），493，494，495，513
right （右），59
elementary divisors （初等因子），155
elementary symmetric functions （初等对称函数），41
elliptic differential operator （椭圆型微分算子），239
equilibrated （均衡的），283
equivalence relation （等价关系）
congruence （相合），221
consimilarity （合相似），251
definition （的定义），45
vector seminorm （向量半范数的），262
equivalent （等价）
matrices （矩阵），164
orthogonally （正交），73
real orthogonally （实正交），73
unitarily （酉），72
vector norms （向量范数），273，279
error analysis （误差分析），335
Euler's theorem （Euler 定理），111
exponential of a matrix （矩阵指数），300
extreme points （端点）
closed convex set （闭凸集），533
doubly stochastic matrices （双随机矩阵），528
extreme ray （极射线）
definition （定义），463
positive semidefinite matrices （半正定矩阵的），164

F

factor analysis （因子分析），431
factorizations （分解），156
Cholesky，114，407
complex skew-symmetric matrix （复斜对称矩阵），217
complex symmetric matrix （复对称矩阵），204
LU，158-165
polar （极），156，411，412ff
product of two Hermitian matrices （两个 Hermite 矩阵之乘积），172
QR，112，164，406

singular value decomposition (奇异值), 411
Takagi, 250, 423, 166
triangular (三角), 157
family ((矩阵)族)
commuting (交换), 51, 81, 99, 139
commuting real normal (实正规交换), 108, 112
complex symmetric (复对称), 243
diagonalizable symmetric (可交换对称), 217
Hermitian (Hermite), 172
normal (正规), 103
simultaneous condiagonalization (同时合对角化), 252
simultaneous diagonalization by congruence (经相合同时对角化), 239
simultaneous diagonalization by unitary congruence (经相合同时对角化), 213
simultaneous singular value decomposition (同时奇异值分解), 126
simultaneous triangularization (同时三角化), 81
Fan (樊畿)
k norms (k 范数), 145
theorem on eigenvalue location (关于特征值位置的定理), 501
Fejer (Fejer)
trace theorem, on positive semidefinite matrices (关于半正定矩阵的迹定理), 459
uniqueness theorem, for elliptic partial differential equations (关于椭圆型偏微分方程的唯一性定理), 460
field (域), 1
of values (值域), 321, 332
forms (型)
bilinear (双线性), 169, 175
Hermitian (Hermite), 174
quadratic (二次), 168, 174, 214, 466
sesquilinear (半双线性), 169
forward substitution (前向替换), 159
fundamental theorem of algebra (代数基本定理), 537

G

general linear group (一般线性群), 14
generalized (广义)
coordinates (坐标), 227
inverse (逆), 421
matrix functions (矩阵函数), 8
Geršgorin
circles (圆), 346
disc theorem (圆盘定理), 344
discs (圆盘), 345, 353
region (区域), 345
Givens's method (Givens 法), 77
Gram-Schmidt process (Gram Schmidt 过程), 15, 148
modified (修改的), 116
symmetric analogue (对称矩阵的类似), 211
graph (图), 168
group (群)
finite Abelian (有限 Abel), 510
general linear (一般线性), 14
isometry (等距), 266, 267
orthogonal (正交), 69, 71
unitary (酉), 69

H

Hadamard
exponential of a matrix (矩阵指数), 161
inequality(不等式), 199, 200, 477, 483
powers of a matrix (幂矩阵), 462
product (乘积), 321, 455, 456, 457, 474, 475
square root of a matrix (矩阵方根), 462
Hahn-Banach theorem (Hahn-Banach 定理), 288
half spaces (半空间), 534
Hermitian
part (部分), 109, 170, 399
property (性质), 260
Hermitian matrices (Hermite 矩阵)
congruent (相合), 223, 224
product of three (三个 Hermite 矩阵的乘积), 469
product of two (两个 Hermite 矩阵的乘积), 172
Hermitian matrix (Hermite 矩阵), 104, 167, 169, 397
analogous to real numbers (比作实数), 170
characterizations (特征), 171

partitioned （分块），175
product with positive definite matrix （与正定矩阵的乘积），465
spectral theorem （的谱定理），171
Hessian （Hessian）167，392，459，534
Hoffman-Wielandt theorem （Hoffman-Wielandt 定理），368，419
homogeneous （齐次性），259，260，290
Hopf's bound （Hopf 界），501
Householder
transformation （变换），71，77，78，117
method （法），78
hyperbolic differential operator （双曲型微分算子），239
hyperplane （超平面），534

I

idempotent （幂等），37，148，311
identity （恒等式）
Newton，44
parallelogram （平行四边形），263
polarization （极化），263
Sylvester，22
ill-conditioned （病态），336
imaginary （虚）
axis （虚轴），532
part of a complex number （复数的虚部），531
inclusion （包含）
principle （原理），189
region （区域），378
independent （无关），3
index （指标）
of an eigenvalue （特征值的），131，139，148
of nilpotence （幂零），37
induced matrix norm （诱导矩阵范数），292
by absolute vector norm （由绝对范数），310
by monotone vector norm （由单调范数），310，365
characterization （特征），302，307
inequality （不等式）
arithmetic geometric mean （算术-几何平均），535
between matrix norms （矩阵范数间的），314
between vector norms （向量范数间的），279
bilinear （双线性），473
Cauchy-Schwarz，15，261，277，535，536
determinant （行列式），351
Fischer，178
Greub and Rheinboldt （Greub 与 Rheinboldt），452
Grunsky，202
Hadamard，199，200，177，183
Hadamard Fischer，185
Holder，276，535
Kantorovich，444，451，452
matrix norm （矩阵范数），290，312
Minkowski，265，536
numerical radius （数值半径），331
Oppenheim，480
Ostrowski Taussky，468，481
positive definite function （正定函数），400
power for numerical radius （关于数值半径的幂），333，334
rank （秩），352
Robertson，468
square root continuity （平方根连续(函数)），411
submultiplicative （次乘性），290
Szasz，479
triangle （三角），259，290
unitarily invariant matrix norm （酉不变矩阵范数），450
unitarily invariant norms （酉不变向量范数），447
Wielandt，442，443
inertia of a matrix （矩阵的惯性），221
infinite series of matrices （矩阵的无穷级数），300
inner product （内积），110
characterization of norm derived from （由内积诱导的范数的特征），263
definition （的定义），260
Frobenius，332
standard （标准），14
usual （普通），14
interior point （内点），282
interlacing （交错）
eigenvalues theorem for bordered matrices （加边矩阵的特征值定理），185
inequalities （不等式），182，185，187，189，

200，104，419
property for singular values （奇异值的性质），419
theorem （定理），182
invariant （不变）
factors （因子），154
subspace （子空间），51
inverse （逆），14
diagonal dominance （对角占优(矩阵)的逆），355
errors in （(矩阵)逆的误差），335
irreducibly diagonally dominant （不可约对角占优(矩阵)的逆），363
minors of （逆的子式），21
series for （逆矩阵的级数），301
small rank adjustment （小秩修正矩阵的逆），18
strict diagonal dominance （严格对角占优(矩阵)的逆），302，349
invertible （可逆），14
irreducible matrix （不可约矩阵），361，362，493，506-515
minimal polynomial criterion （的极小多项式准则），515
irreducible normal form （不可约正规形式），506
irreducibly diagonally dominant （不可约对角占优），362
isometry （等距），68
for a vector norm （关于向量范数），266
isomorphism （同构），4

J

Jacobi
identity （恒等式），21
method （法），76
Jordan
block （块），121
normal form （法式），121
Jordan canonical form （Jordan 标准形），121，129
real （实），152
theorem （定理），126

K

Kakeya's theorem （Kakeya 定理），318
kernel （核），456，162
Kojima's bound on zeroes （零点的 Kojima 界），319，364
Krein Milman theorem （Krein Milman 定理），533，534
Kronecker product （Kronecker 乘积），474，475
Krylov sequence （Krylov 序列），164

L

Lagrange
equations （方程组），227
interpolating polynomial （插值多项式），29，188，405
interpolation （插值法），29
interpolation formula （插值公式），30
Lanczos tridiagonalization （Lanczos 三对角化，164
Laplace
equation （方程），239
expansion （展开式），7
least squares （最小二乘）
approximation （逼近），429，431，515
solution （解），421
left Perron vector （左 Perron 向量），497
Levy Desplanques
condition for invertibility （可逆性条件），302，349
theorem （定理），302，349
limit （极限）
of a sequence （序列的），270
point （点），282
line segment(线段)，289
linear(线性)
dependence(相关)，3
independence(无关)，3，407
transformation(变换)，5
loop(圈)，358

M

majorization （优化），199，425，446
and unitarily invariant norms （与酉不变范数），447
characterizations （特征），197
definition （定义），192
eigenvalues by diagona. entries （(方阵)对角元组

成的向量优化(其)特征向量), 193, 196
product inequality (乘积不等式), 199
spectrum of a sum ((矩阵)和的谱), 194
Markov chain(Markov 链), 497
Mason and Carmichael's bound on zeroes(零点的 Mason 和 Carmichael 界), 317, 318, 364
matrix (矩阵)
adjacency (邻接), 168, 523
almost diagonalizable (几乎可对角化), 89
approximation problems (逼近问题), 427
backward identity (后向单位), 28, 207
block diagonal (分块对角), 24
block triangular (分块三角), 25, 90
bordered (加边), 185
change of basis (基变换), 32
circulant (轮换), 26
combinatorially symmetric (组合对称), 523
commuting (交换), 135
companion (友), 147, 149, 316
complex orthogonal (复正交)71, 72
complex symmetric (复对称), 201
compound (复合), 19
convergent (收敛), 137
correlation (相关), 400
covariance (协方差), 219, 239, 392, 424
diagonal (对角), 23
diagonalizable (可对角化), 46, 139, 145
doubly stochastic (双随机), 197, 527 529
equivalent (等价), 164
essentially nonnegative (本性非负), 506
essentially triangular (本性三角), 26
function of a (函数), 300
Gram, 407
Hankel, 27, 202, 393
Hermitian, 109, 167, 169
Hessenberg, 28
Hessian, 392, 459, 534
Hilbert, 341, 401
identity (单位), 6
indefinite (不定), 397
indicator (指标), 356
irreducible (不可约), 361, 362
Jacobian, 218
Jordan, 121, 129
negative definite (负定), 397
nilpotent (幂零), 139
nonderogatory (非减次), 135
normal (正规), 100
normal skew-symmetric (正规斜对称), 217
normal symmetric (正规对称), 207
orthogonal (正交), 71, 72
orthogonally diagonalizable (正交对角化), 211
orthostochastic (正交随机), 197
permutation (置换), 25
rank one (秩 1), 61
real orthogonal (实正交), 66, 72, 107
real skew-symmetric (实斜对称), 107
real symmetric (实对称), 107
reducible (可约), 360
scalar (纯量), 23
similarity (相似), 14
skew-Hermitian (斜 Hermite), 100, 169
skew-orthogonal (斜正交), 72
skew-symmetric (斜对称), 109
skew-symmetric normal (斜对称正规), 217
stochastic (随机), 526 529
symmetric (对称), 167, 201
symmetric diagonalizable (对称可对角化), 211
symmetric normal (对称正规), 207
symmetric unitary (对称酉), 215
Toeplitz, 27, 136, 137, 394, 409, 456, 462
triangular (三角), 24
tridiagonal (三对角), 28, 174, 395, 409, 506
unitary (酉), 66, 109
unitary characterizations (的酉特征), 67
unitary symmetric (酉对称), 215
Vandermonde, 29
weakly irreducible (弱不可约), 383
matrix functions, generalized ((广义)矩阵函数), 8
matrix norm (矩阵范数)
bound norm (有界范数), 294
Euclidean norm (Euclid 范数), 291
Frobenius norm (Frobenius 范数), 291
generalized (广义), 320 342

Hilbert-Schmidt norm (Hilbert Schmidt 范数), 291
induced (诱导), 307, 365
induced by a similarity (由相似诱导的), 296
induced by vector norm (由向量范数诱导的), 292, 294
inequality for (关于不等式), 312
l_1 norm (l_1 范数), 291
l_2 norm (l_2 范数), 291
nl norm (nl 范数), 292
lub norm (lub 范数), 294
maximum column sum norm (极大列和范数), 294
maximum row sum norm (极大行和范数), 295
minimal (极小), 306, 307
not convex set (的非凸集), 312
operator norm (算子范数), 294
Schatten p norm (Schatten$p$ 范数), 441
Schur norm (Schur 范数), 291
self adjoint (自伴), 309
spectral norm (谱范数), 295, 441
trace norm (迹范数), 44
unitarily invariant (酉不变), 292, 296, 308
max min theorem (极大极小定理), 179, 493, 496, 504
maximal element (极大元), 381
maximum of a continuous function (连续函数的极大值), 541
McCoy's theorem (McCoy 定理), 94, 97
Mercer's theorem (Mercer 定理), 456
metric convex hull (度量凸包), 289
min-max theorem (极小-极大定理), 179, 493, 496
minimal polynomial (极小多项式), 142, 145
algorithm to compute (的算法), 144, 148
definition (的定义), 143
diagonalization criterion (对角化的极小多项式准则), 145
of a direct sum (直和的), 149
minimally spectrally dominant (极小谱优势), 330
Minkowski determinant inequality (Minkowski 行列式不等式), 482
minor (子式), 17
principal (主), 17, 40, 398
signed (带正负号), 17
moments of eigenvalues (特征值的矩), 43, 44
moment sequence (矩序列)
Hausdorff, 393
Toeplitz, 393
moments (矩)
algebraic (代数), 393
trigonometric (三角), 393, 455, 456
monotonicity theorem (单调性定理), 182
Montel's bound on zeroes ((多项式)零点的 Montel 界), 317, 318, 364
Moore Penrose generalized inverse (Moore Penrose 广义逆), 421
multilinear (多重线性), 11

N

Nehari's theorem (Nehari 定理), 202
nodes (结点), 168
nondefective (非亏损), 58, 103
nonderogatory (非减次), 58, 135, 147
nonnegative (非负), 259, 260, 290
nonnegative matrix (非负矩阵), 359
applications (的应用), 487-490
definition (的定义), 490
doubly stochastic (双随机), 527-529
eigenvalues (特征值), 489, 503 505, 507-515
eigenvectors (特征向量), 489, 503-505, 507-515
general limit theorem (一般极限定理), 524
irreducible (不可约), 507 515
limit of powers (幂的极限), 489
limit theorems (极限定理), 500, 516, 524, 525
Perron root (Perron 根), 505, 508
Perron vector (Perron 向量), 505, 508
Perron-Frobenius theorems (Perron-Frobenius 定理), 508-511
primitive matrices (素矩阵), 515-524
spectral radius (谱半径), 489, 491-495, 503-505, 507-515
stochastic (随机), 526-529
nonsingularity, characterizations of (非奇异性的特征), 14
nontrivial cycle (非平凡回路), 383

norm （范数）
  and inversion （与可逆性），301
  characterization of derived （诱导范数的特征），263
  compatible （相容），294，324
  consistent （协和），324
  dual （对偶），275，410
  dual pair （对偶对），278
  minimally spectrally dominant （极小谱优势），330
  pre norm （准范数），272
  spectrally dominant （谱优势），324
  subordinate （从属），324
normal matrix （正规矩阵）
  characterizations （特征），101，108-112
  definition （定义），100
  perfectly conditioned eigenvalues （优态特征值），367
  real （实），104ff
null space （零空间），5，262
numerical radius （数值半径），321，331，332，333，334
numerical range （数值范围），321

O

open set （开集），282，541
operator norm （算子范数），291
orthogonal （正交）
  complement （分量），16
  vectors （向量），15
orthogonality （正交性），15
orthogonally （正交）
  diagonalizable （可对角化），101
  equivalent （等价），73
orthonormal （标准正交），15
  basis （基），16
Ostrowski
  condition for invertibility （可逆性条件），381
  region （区域），378，379
  theorem （定理），224，378
ovals of Cassini （Cassini 椭圆形），380

P

partitioned matrix （分块矩阵）
  definition （的定义），17
  inverse （的逆），18，472
  Schur complement，472
Pearcy's theorem （Pearcy 定理），76
perfectly conditioned （优态），336
permanent （积和式），8
permutation invariant （置换不变(函数)），368，438
Perron
  root （根），497，505，508
  theorem （定理），500
  vector （向量），497，505，508
Perron-Frobenius theorems （Perron-Frobenius 定理），508-511
perturbation （扰动）
  eigenvalues （特征值），198，343，364
  linear equations （线性方程组），335
  theorems （定理），364
plane rotations （平面旋转），74
Poincaré separation theorem （Poincaré 分离定理），190，441
polar （极）
  coordinates in complex plane （复平面的极坐标），532
  decomposition （极分解）
polar form （极形式），156，411，412ff
  examples and applications （例及应用），427ff
polynomial （多项式）
  bounds for zeroes of （零点的界），316-319
  continuous dependence of zeroes on coefficients （的零点对(其)系数的连续依赖），539
  for inverse （逆矩阵的），88
  in a matrix （矩阵的），36，135，142
  monic （首一），142
  similarity invariants （相似不变量），154
  zeroes of a （的零点），537
  zeroes of a real （实的零点），538
poorly conditioned （坏条件），336
positive （正(定)），259，260，290
  cone （锥），398
  definite function （函数），400，401，463
  definite kernel （核），402，462
positive definite matrix （正定矩阵），250

applications （的应用），391-396，459
characteristic polynomial （的特征多项式），403
characterizations （的特征），402
concavity of logarithm of the determinant （行列式的对数凹性），466
concavity of trace of the inverse （的逆的迹的凹性），468
definition （定义），396
determinant criterion （行列式判别准则），404
determinantal inequalities （行列式不等式），476-486
eigenvalues （的特征值），402
kth root （k 次根），405
ordering （次序关系），469ff
positive matrix （正矩阵），359
definition （的定义），490
eigenvalues （特征值），495-503
eigenvectors （特征向量），495-503
left Perron vector （左 Perron 向量），497
Perron root （Perron 根），497
Perron vector （Perron 向量），497
Perron's theorem （Perron 定理），500
spectral radius （谱半径），495-503
positive semidefinite matrix （半正定矩阵），182
definition （定义），396
kth root （k 次根），405
ordering （次序关系），469ff
rank k （秩 k），457
positive semidefinite ordering （半正定次序关系），469
power method （幂法），62，523
preorder （预序），384
primitive matrix （素(或本原)矩阵），515-524
characterizations （的特征），516，517
definition （的定义），516
directed graph （的有向图），517
eigenvalues （的特征值），516
Holloday-Varga theorem （的 Holloday-Varga 定理），520
index of primitivity （的本原指标），519
Wielandt's theorem （Wielandt 定理），520
procrustean transformation （强行变换），431
property （性质）
L，97，100
P，100
SC，355，358，359，362
pure imaginary complex number （纯虚复数），532

Q

QR
algorithm （算法），114
factorization （分解），112，164，406
quasi-linearization （拟线性化），191，453，455，486

R

range （值域），5
rank （秩），12
equalities （秩的等式），13
inequalities （秩的不等式），13，175，352，458
rank one （秩 1）
approximation （逼近），427
limit （极限），499
matrix(矩阵)，61
rational(有理)
canonical form(标准形)，156
form （形），154
Rayleigh Ritz
ratio （比），176
theorem （定理），176，422
real （实）
axis （实轴），532
Jordan canonical form （实 Jordan 标准形），152
part of a complex number （复数的实部），531
rectangular coordinates in complex plane （复平面中的直角坐标），532
residual vector （剩余向量），338，373，374
reverse order law （倒序律），6
Riemann sum （Riemann 和），462
right （右）
eigenvector （右特征向量），59
half-plane （右半平面），532
Romanovsky's theorem （Romanovsky 定理），517
rotation problem （旋转问题），431，435
round-off errors （舍入误差），335

row rank （行秩），12
row-reduced echelon form（RREF）（行简化梯形阵），10

S

scalar product （纯量积），14
Schur
complement （补），21，472
majorization theorem （优化定理），193
norm （范数），291
product （乘积）
product theorem （乘积定理），455，458
unitary triangularization theorem （酉三角化定理），79，83
self-adjoint vector norm on matrices （关于矩阵的自伴向量范数），450
sensitivity （灵敏度）
of eigenvalues （特征值的），372
of eigenvectors （特征向量的），373
of solutions of linear equations （线性方程组解的），339
separating hyperplane theorem （分离超平面定理），534
shift operator （移位算子），38
signature of a matrix （矩阵的符号差），221
signum(sgn) （正负号），8
similarity （相似(性)），44
inverse to adjoint （逆相似于伴随），70
matrix and its transpose （矩阵与其转置），134
to a real matrix （相似于实矩阵），172
to a real matrix，$A\bar{A}$ （$A\bar{A}$ 相似于实矩阵），253
to adjoint （相似于伴随），172
simultaneous （同时）
*congruence，canonical pairs （*相合下的标准形偶），236
condiagonalization （合对角化），252
simultaneous diagonalization （同时对角化），49，52
by *congruence （用*相合），240，464ff
by congruence （用相合），228，241，250
by unitary congruence （用酉相合），228，235
characterization of （的特征），228
singular （奇异），14
singular value decomposition （奇异值分解），157，205，325，411，414ff，421
examples and applications （的例及应用），427ff
simultaneous （同时），426
singular values （奇异值），205，415
largest （最大），421
of a product （乘积的），423
of a sum （和的），423
perfectly conditioned （优态），418
variational characterization （的变分特征），420
skew Hermitian （斜 Hermite）
matrix （矩阵），100，169
part （部分），109，170，399
solution equivalent systems （解等价方程组），11
span （张成）2，3
Specht's theorem （Specht 定理），76
spectral characteristic （谱示性数），330
spectral radius （谱半径），35，198，296，313，348，489
as a limit of matrix norms （作为矩阵范数极限的），297
as a limit using norms or pre norms，（作为用范数或准范数表示的极限的），299，322
spectral theorem （谱定理）
Hermitian matrices （Hermite 矩阵的），104，171
normal matrices （正规矩阵的），101，425
spectrally dominant （谱优势），329
spectrum （谱），35
of a sum （(矩阵)和的），92
of a sum by majorization （用优化表示和的谱），194
square root of a matrix （矩阵的平方根），54，405
continuity of （的连续性），411
standard （标准）
basis （基），4
inner product （内积），14
strictly diagonally dominant （严格对角占优(矩阵)），302，349
strongly connected directed graph （强连通有向图），358，362，383
submatrix （子矩阵），4
eigenvalues （的特征值），189
principal （主），17，397

submultiplicative (次乘性), 290
subspace (子空间), 2
invariant (不变), 51
Sylvester's law of inertia (Sylvester 惯性定律), 223, 238
analogue for symmetric matrices (对称矩阵的模拟), 225
homotopy proof (同伦证明), 242
symmetric (对称)
gauge function (度规函数), 438, 445
symmetric matrix (对称矩阵), 167, 397
diagonalizable (对角化), 211
every matrix similar to a (每个矩阵相似于一个), 209
product of two (两个对称矩阵的乘积), 210
real (实), 169, 218
Szasz's inequality(Szasz 不等式), 479

T

Takagi's factorization (Takagi 分解), 204, 423, 466
as consimilarity analogue of spectral theorem (类似于谱定理的合相似下的), 250
Hua's proof (Hua 证明), 217
Siegel's proof (Siegel 证明), 216
Taussky's theorem (Taussky 定理), 363
topological notions (拓扑概念), 282, 288
trace (迹), 40, 175, 398
norm (迹范数), 441, 445
zero (迹为零), 77
transpose (转置(矩阵)), 6
transposition (对换(矩阵)), 25
triangularization (三角化)
by consimilarity (用合相似), 244, 245
by unitary congruence (用酉相合), 203
by unitary consimilarity (用酉合相似), 244, 245
by unitary similarity (用酉相似), 79
orthogonal (正交), 84
simultaneous (同时), 81, 84, 94
tripotent (三次幂等), 148
trivial cycle (平凡回路), 358
truncation errors (截断误差), 335

U

unit (单位)
disc (圆盘), 532
sphere (球面), 273
unit ball (单位球), 273, 281
compact (紧), 283, 284
convex (凸), 284
equilibrated (均衡), 284
properties of (的性质), 283
unitarily diagonalizable (酉可对角化), 101
unitarily equivalent (酉等价)
definition (定义), 72
equal diagonal entries (对角元相等的矩阵), 77
Specht's theorem (酉等价的 Specht 定理), 76
to upper triangular matrix (于上三角矩阵), 79
unitarily invariant matrix norms (酉不变矩阵范数), 296, 308
set is convex (之集是凸集), 450
unitarily invariant vector norm, on matrices (关于矩阵的酉不变向量范数), 437, 441, 445
Von Neumann's characterization (的 Von Neuman 特征), 438
when a matrix norm (何时是矩阵范数), 450
unitary group (酉群)
unitary matrix (酉矩阵), 66-72, 157
characterizations (的特征), 67
definition (的定义), 66
selection principle (的选择原理), 69
upper half-plane (上半平面), 532

V

variational characterization of eigenvalues (特征值的变分特征), 176
vector (向量)
normalized (正规化), 15
unit (单位), 15
vector norm (向量范数)
absolute (绝对), 285, 310, 365, 438
algebraic properties (代数性质), 268
analytic properties (分析性质), 269
Cauchy sequence with respect to a (关于向量范数的 Cauchy 序列), 274
characterization via unit ball (的单位球的特征), 284

compatible （相容），324，327
consistent （协和），321
convergence with respect to *a* （关于向量范数收敛），269
definition （的定义），259
derived from inner product （由内积诱导的），262
dual （对偶），275
duality theorem （对偶性定理），287
equivalent （的等价），273，279
Euclidean （Euclid），264
geometric properties （几何性质），281
l_1 norm （l_1 范数），265
l_2 norm （l_2 范数），264
l_p norm （l_p 范数），265
l_∞ norm （l_∞ 范数），265，322
L_1 norm （L_1 范数），266
L_2 norm （L_2 范数），266
L_p norm （L_p 范数），267
L_∞ norm （L_∞ 范数），267
Manhattan norm （Manhattan 范数），265
max norm （极大范数），265
monotone （单调），285，310，365，449
on matrices （关于矩阵的），320-342
polyhedral （多面），282
pre norm （准范数），272，322
spectrally dominant （谱优势），329
subordinate （从属），324
sum norm （和），265
uniformly continuous （一致连续(性)），271
unit ball （单位球），273，281
unit sphere （单位球面），273
unitarily invariant （酉不变），265，267
unitarily invariant on matrices （矩阵的酉不变），437
weakly monotone （弱单调），285
vector seminorm （向量半范数），259
vector space （向量空间），2
complete （完备），274
Von Neumann
inner product theorem （内积定理），263
unitarily invariant norm theorem （酉不变范数定理），438

W

weak minimum principle （弱极小原理），460
weakly connected directed graph （弱连通有向图），383
Weierstrass's theorem （Weierstrass 定理），541
weighted arithmetic-geometric mean inequality （加权算术-几何平均值不等式），535
well conditioned （良态的），336
Weyl's theorem （Weyl 定理），181，184，367，419，423
Wielandt
example （例），522
theorem （定理），520
Witt cancellation theorem （Witt 消去定理），78，141